

Chromospheric Variability: Analysis of 36 years of Time Series from
the National Solar Observatory/Sacramento Peak
Ca II K-line Monitoring Program

(Short title: Solar Ca II K-line Variability)

Jeffrey D. Scargle

Space Science and Astrobiology Division, MS 245-3

NASA Ames Research Center

Moffett Field, CA 94035-1000

`jeffrey.d.scargle@nasa.gov`

Stephen L. Keil,

National Solar Observatory

`Keil@nso.edu`

Simon P. Worden

NASA Ames Research Center

`pete.worden@nasa.gov`

ABSTRACT

Analysis of more than 36 years of time series of seven parameters measured in the NSO/AFRL/Sac Peak K-line monitoring program elucidates five components of the variation: (1) the solar cycle (period ~ 11 years), (2) quasi-periodic variations (periods ~ 100 days), (3) a broad band stochastic process (wide range of periods), (4) rotational modulation, and (5) random observational errors. Correlation and power spectrum analyses elucidate periodic and aperiodic variation of the chromospheric parameters. Time-frequency analysis illuminates periodic and quasi periodic signals, details of frequency modulation due to differential rotation, and in particular elucidates the rather complex harmonic structure (1) and (2) at time scales in the range $\sim 0.1 - 10$ years. These results using only full-disk data further suggest that similar analyses will be useful at detecting and characterizing differential rotation in stars from stellar light-curves such as those being produced by NASA's Kepler observatory. Component (3) consists of variations over a range of timescales, in the manner of a $1/f$ random noise process. A time-dependent Wilson-Bappu effect appears to be present in the solar cycle variations (1), but not in the stochastic process (3). Component (4) characterizes differential rotation of the active regions, and (5) is of course not characteristic of solar variability, but the fact that the observational errors are quite small greatly facilitates the analysis of the other components. The recent data suggest that the current cycle is starting late and may be relatively weak. The data

analyzed in this paper can be found at the National Solar Observatory web site http://nsosp.nso.edu/cak_mon/, or by file transfer protocol at <ftp://ftp.nso.edu/idl/cak.parameters>.

Subject headings: Solar Physics

Contents

1	The K-line Monitoring Program	5
2	The Time Series	7
3	Autocorrelation Analysis	17
4	Power Spectrum Analysis	19
5	Time-Frequency Distributions	21
6	Cross- Analysis	24
7	Wilson-Bappu Effect	28
8	Conclusions	30
9	Appendix 1: Power Spectra and Time-Frequency Distributions	33
10	Appendix 2: Notes on the Computations	41

1. The K-line Monitoring Program

For nearly four decades the NSO/AFRL/Sac Peak K-line monitoring program (Keil and Worden 1984) has produced almost daily measurements of seven parameters characterizing the Ca II K-line integrated over the solar disk. This continuing program is aimed at characterizing chromospheric variability due to various processes and on various time-scales. Beginning on November 20, 1976 the time series now cover more than three 11-year solar cycles (21, 22, and 23) or $1\frac{1}{2}$ Hale Cycles. Now is a good time to analyze these data, allowing comparison of two alternate cycles as well as providing preliminary information about the beginning of the current cycle 24. This paper describes data up to December 20, 2012.

The motivation for this survey and details of observational procedures are given in (Keil and Worden 1984; Keil et al. 1998; White et al. 1998), and further documentation and data are on the NSO website (Keil et al. 2011). Table 1 of (Keil et al. 1998) describes the seven measured K-line parameters. These parameters, in the order and with the boldface tokens as they appear in the data file posted at http://nsosp.nso.edu/data/cak_mon.html, are:

- 1.**EMDX**: Emission Index, equivalent width in 1 Å band centered on the K-line profile
- 2.**VIORED** : $I(K_{2V})/I(K_{2R}) = [I(K_{2V}) - I(K_3)]/[I(K_{2R}) - I(K_3)]$, ratio of blue to red emission maxima
- 3.**K2VK3** : $I(K_{2V})/I(K_3)$, strength of blue wing relative to K_3
- 4.**DELK1** : $\lambda(K_{2R}) - \lambda(K_{2V})$, wavelength separation of the two emission maxima
- 5.**DELK2** : $\lambda(K_{1R}) - \lambda(K_{1V})$, wavelength separation of the two emission minima
- 6.**DELWB** : Wilson-Babbu parameter, wavelength separation of outer emission maxima edges
- 7.**K3** : K_3 , intensity in the core of the K-line

Further description of the parameters is as follows, quoted (with reordering) from the NSO web site:

Several K-line parameters, including the emission index and various measures of asymmetry, are abstracted from the calibrated K-line profiles and stored on the NSO ftp site. These parameters include: (1) the Ca K-line emission index which is defined as the equivalent width of a 1 angstrom band centered on the K-line core, (2) the K-line asymmetry which is the ratio of the blue and red K2 emission maxima ($K2V/K2R$), (3) the relative strength of the blue K2 emission peak with respect to the K3 intensity ($K2V/K3$), (4) the separation of the two emission maxima ($K2V-K2R$), (5) the separation of the blue and red K1 minima ($K1V-K1R$), (6) the Wilson-Bappu parameter which is the width measured between the outer edges of the K2 emission peaks, and (7) the K3 intensity (the core intensity).

The schematic K-line profiles in (Keil and Worden 1984) and in Fig. 1 of (Donahue and Keil 1995) clarify these definitions. Note that 1 and 7 are K-line intensities, expressed as an equivalent width and a percentage of the continuum, respectively; 2 and 3 are intensity ratios, and 4, 5 and 6 are wavelength separations of K-line features in Angstroms.

Figure 1 shows the number of days on which observations have been obtained during 30 day intervals. It is an update of Figure 1 of (Keil et al. 1998) in the same format but with slightly different interval boundaries. If the observation times are independent random variables with a changing rate (also known as a variable-rate Poisson process) the blocks in the figure are statistically the best step-function representation of the variation of the event rate, obtained using the Bayesian Blocks algorithm (Scargle 1998; Jackson *et al.* 2005; Scargle et al. 2013). The mean and median interval between samples is 3.39 days and 1 day, respectively.

We here report exploratory analysis of these time series data, aimed at characterizing the variability of the individual parameters and possible relationships between them. A related analysis Integrated Sunlight Spectrometer of the National Solar Observatory is given by (Bertello et al. 2012). For background the reader may consult the review paper (Hall 2008) on stellar chromosphere activity and the book (Schrijver and Zwaan 2000) provides an excellent overview of the relevant solar and stellar physics. The paper (Livingston et al. 2007) describes analysis of McMath Solar Telescope data similar to those described here.

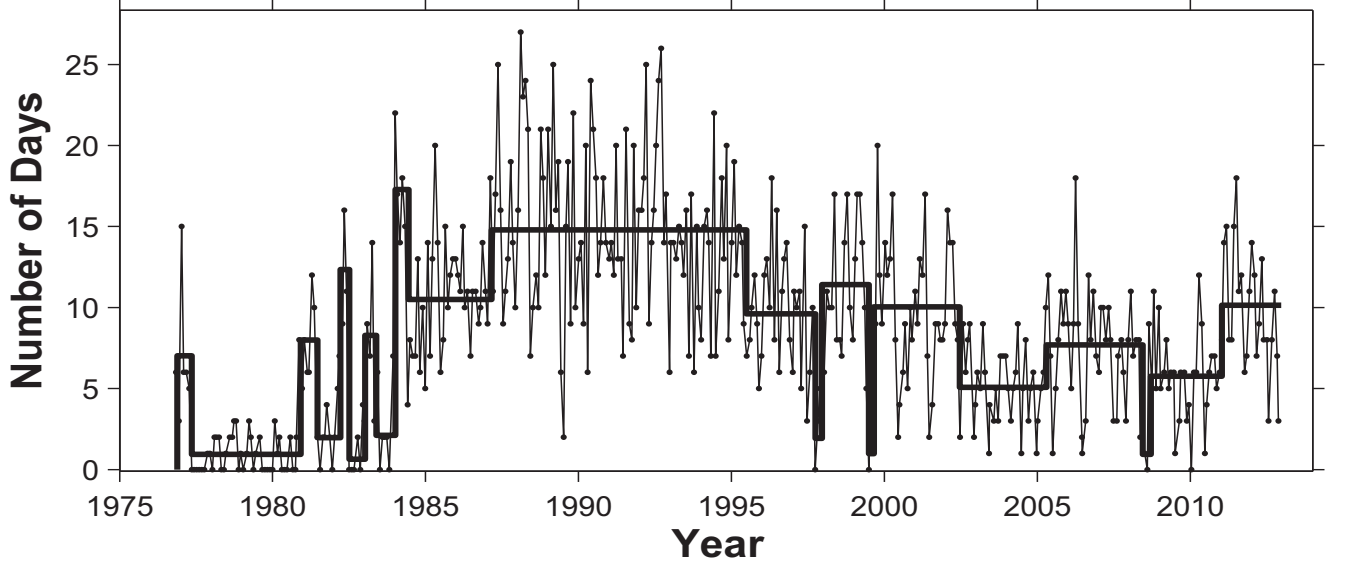


Fig. 1.— Number of Days Observed per 30 Day Interval. The heavy lines denote Bayesian blocks, indicating statistically significant changes in the sampling rate.

The following sections describe time domain, correlation, power spectrum, and time-frequency analyses carried out on these data. The emission index **EMDX** and core intensity **K3** are emphasized, because these two closely related parameters vary in quite similar ways and seem to be the most straightforward diagnostics of chromospheric activity. No data preprocessing beyond that described in (Keil and Worden 1984) was applied, other than the removal of a few outliers.

2. The Time Series

Figure 2 presents these 3894 observations up to December 20, 2012 in the same format as in Figure 2 of (Keil et al. 1998) with the exception that the order is the same as listed above and a few outlying points presumed to be erroneous have been removed. An estimate of an upper limit to the 1σ observational error variance is plotted as a vertical bar near the bottom of each panel above the date 1980. These errors were determined from the auto-correlation function of the time series data, as described in §3. Note that these upper limits on the errors are quite small; even the apparently random variations are mostly real.

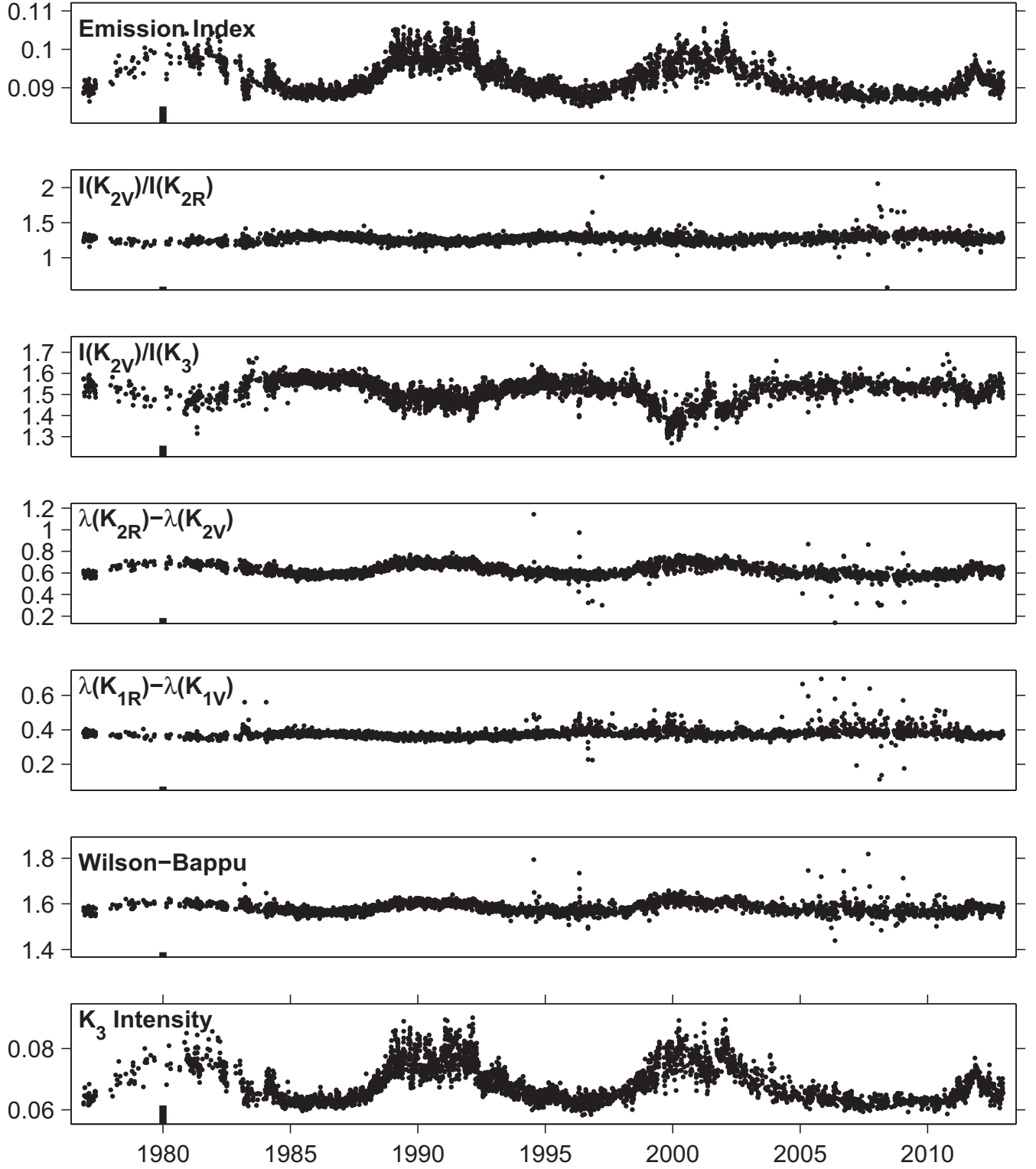


Fig. 2.— Ca II K-line time series data, with a few outliers not plotted. Just above the date 1980 is a small bar representing the upper limit on the average error variance determined directly from the data as described in §3.

Figure 3 shows an enlarged version of the two intensity time series, **EMDX** and **K3**. The lines are fits using the MatLabTM (The Mathworks, Inc.) spline function `spaps`. The resulting smoothing has the effect of removing or reducing the shorter time-scale components, thus elucidating time-scales longer than roughly a fraction of a year. Shown are fits with two different values for the spline’s error tolerance parameter. Roughly speaking the lesser smoothing reveals the solar cycle and the somewhat faster quasi-periodic variations to be discussed below, while the greater smoothing mostly removes the latter and emphasizes the former. These degrees of smoothing correspond to two nearly equal maxima of a measure of independence between the fit and the corresponding residuals (data minus fit).

In these variables there is indication of a relative weakness and lateness for the developing cycle 24, as emphasized in Figure 4. It will be interesting to see whether the rather sharp peak over the last few years will be followed by more up and down activity or whether it marks the end of the cycle. In the NSO helioseismic data Cycle 24 also seems to be late in showing up at high latitudes. The ratio of small sunspots to large appears to have increased in this cycle, perhaps accounting for the decrease in sunspot magnetic fields suggested by earlier observations, e.g. (Livingston et al. 2012).

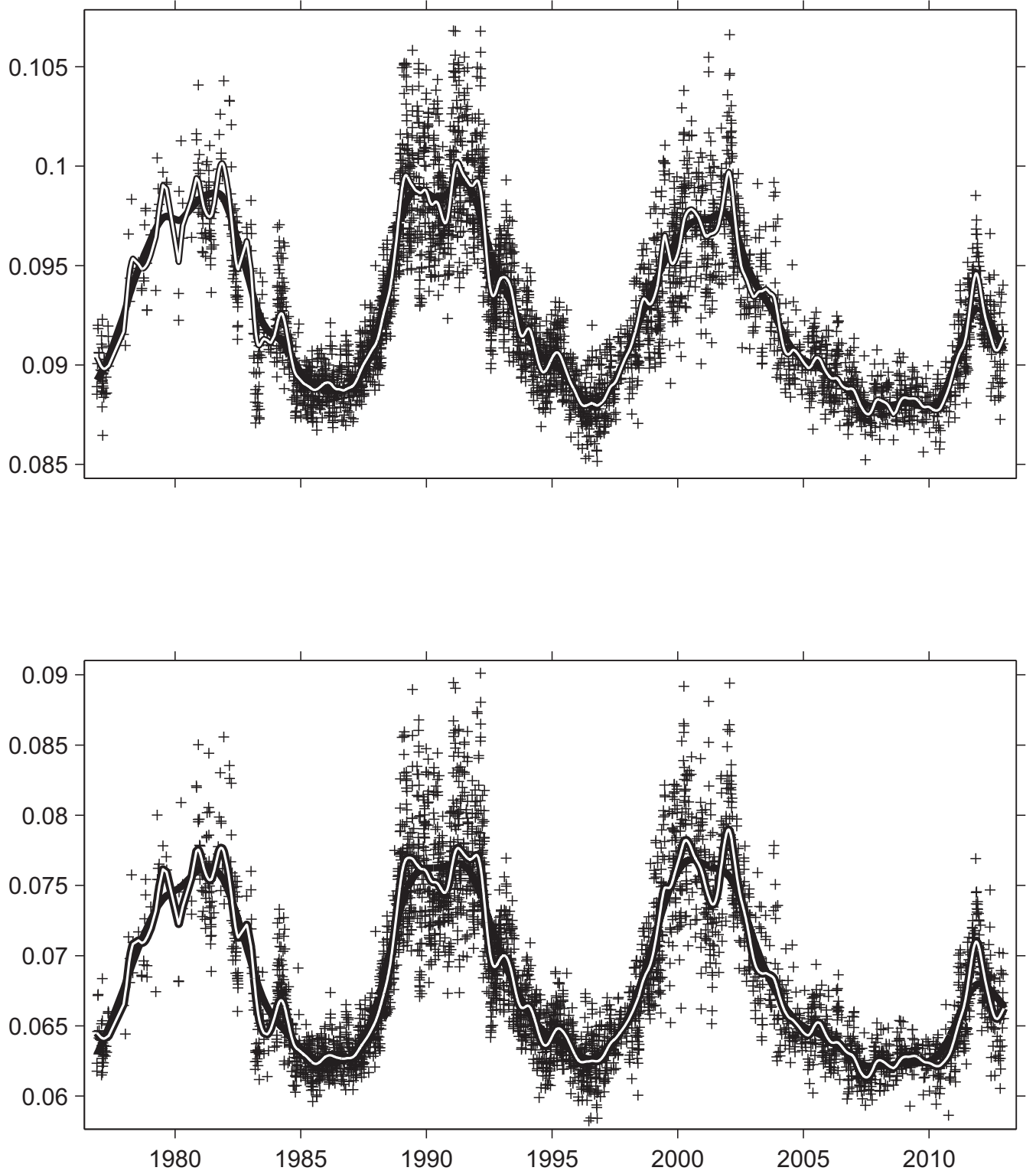


Fig. 3.— Time series of the **EMDX** (top) and **K3** (bottom), with spline fits computed with two different degrees of smoothing: greater smoothing as the dark line, less smoothing as the white-filled line.

Figure 4 makes a side-by-side comparison of the two variables **EMDX** and **K3**, with both degrees of smoothing. Figures 3 and 4 between them make several points: (1) these two intensity variables, under either of the adopted smoothing choices, have very similar behavior; (2) the more complex structure corresponding to the smaller degree of smoothing is very similar over the three cycles; (3) the onset of cycle 24 follows an inter-cycle interval that, compared to the previous two, has been relatively long and quiescent, and (4) this onset is relatively late and may also be relatively quiescent. The first point is expected because these variables measure similar aspects of the central depth of the K-line. The second is perhaps surprising, as it suggests that the cycle of chromospheric activity is repeatably more complex than a series of simple monotonic rises to maximum followed by decline to minimum. We regard the repeatability of the irregular structures in the plots in the bottom panel as evidence that they correctly represent the consistent behavior of these parameters over the solar cycle. Note especially these common features of cycles 21 and 23: three sharp peaks near solar maximum, with similar peaks on both the rising and falling parts of each cycle. (The agreement between the structure for the two independent parameters **EMDX** and **K3** is further evidence for the relative unimportance of random observational errors.) As a counterpoint to this consistency the current cycle is already different from the previous three – items (3) and (4) – but of course a full assessment cannot be made for several more years.

There are four types of variability, plus observational errors, present in all of the time series: (1) a trend obviously tied to the 11-year solar cycle, (2) quasi-periodic signals an order of magnitude faster than (1), (3) random flicker noise, (4) a periodic signal at or about the solar surface rotation frequency, and (5) the inevitable errors of observation. The first two of these are the relatively smooth variations just discussed. The third and fourth are difficult to distinguish from each other visually in light curves. However the major part of the variability in the magnified plot in Figure 5 is rotational modulation. While there is not a precise one-to-one correspondence between peaks and the fiducial lines at the solar rotation cadence, the presence of a periodicity with an amplitude well above the observational errors is strongly suggested. In §4 and §5 (2), (3) and (4) are clearly separated using power spectrum and time-frequency analysis of the residuals obtained by subtracting a smoothed fit from the raw data.

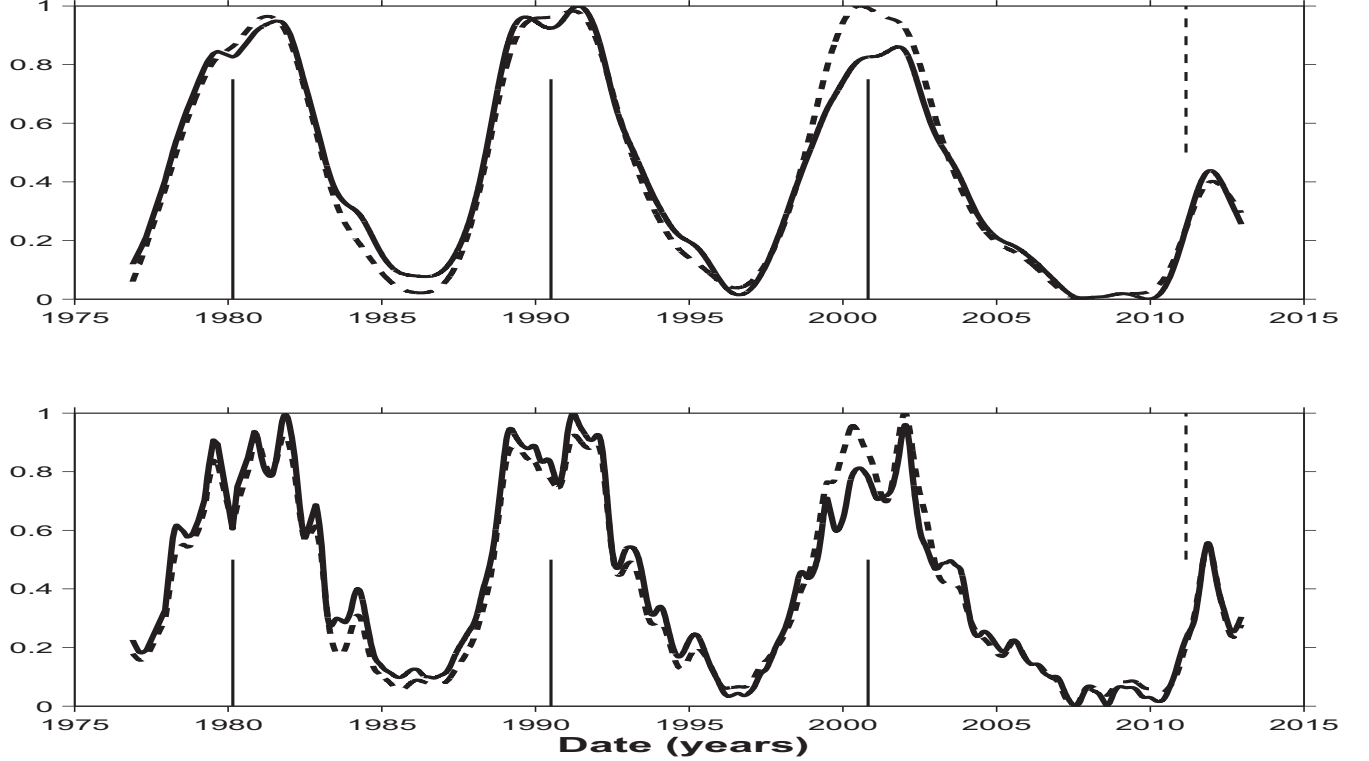


Fig. 4.— Spline fits to **EMDX** = solid line and **K3** = dashed line (with outliers removed) with more (top) and less (bottom) smoothing. Vertical lines indicate times of maxima obtained by fitting a polynomial of degree 3 to the smoothed **EMDX** light curve in the bottom panel. A dashed line indicates when the next maximum would have occurred if the mean interval of 10.5 years between these three maxima had continued.

While this phenomenological separation may not mean that there really are four independent physical processes, all of them are clearly real and originate from chromospheric activity, or a modulation thereof in case (4). The observational errors are small, as demonstrated in §3. In addition the details of the variation of **EMDX** and **K3** are much the same (see Figure 4), which would not be the case if observational errors were significantly large. It is difficult *a priori* to rigorously identify the physical processes underlying these components, but the properties listed in Table 1 argue for distinct physical origins of the components.

A positive amplitude-variance correlation is clearly evident in the **EMDX** and **K3** time series, and less prominently the others: variance large near the peak and small near the

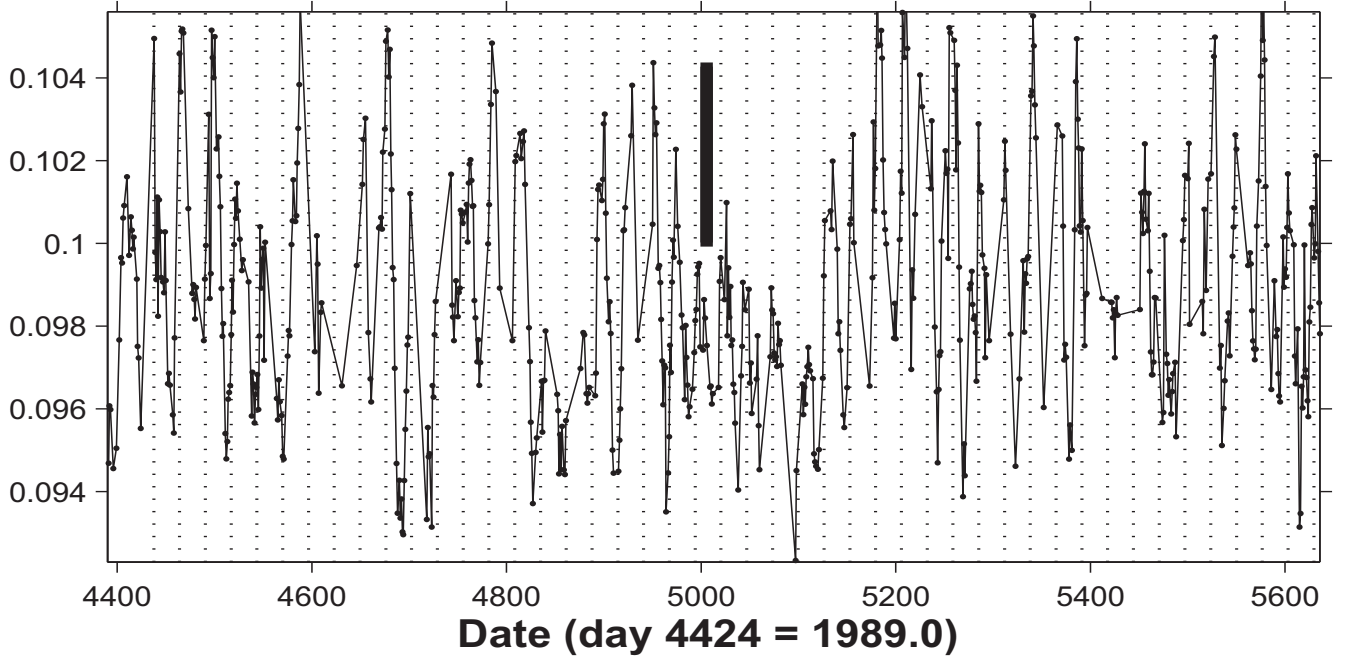


Fig. 5.— A short segment of the **EMDX** light-curve near the peak of cycle 22. The vertical dotted lines are separated by the period corresponding to the peak of the power spectrum of the data in this 3.4-year long interval. (Accordingly this period, 26.49 days, has meaning only for this limited time window.) The solid bar in the middle/top part of the figure is the 1σ observational error variance determined in §3.

valleys. The plot of the residuals from a smooth fit in Figure 6 makes this effect even more obvious. In view of the large contribution to the variance from rotationally modulated chromospheric activity (*cf.* Figure 5) closely following the solar cycle, such correlations are expected.

Table 1: Five Modes of Variability

	Amplitude	Time-Scale	Nature
(1) Solar cycle	large	long (≈ 11 years)	deterministic
(2) Reiger-type periods	small	medium (≈ 100 days)	quasi-periodic
(3) Flicker noise	small	large range	random
(4) Rotation modulation	medium	short (27 days)	periodic
(5) Observation errors	small	instantaneous	random

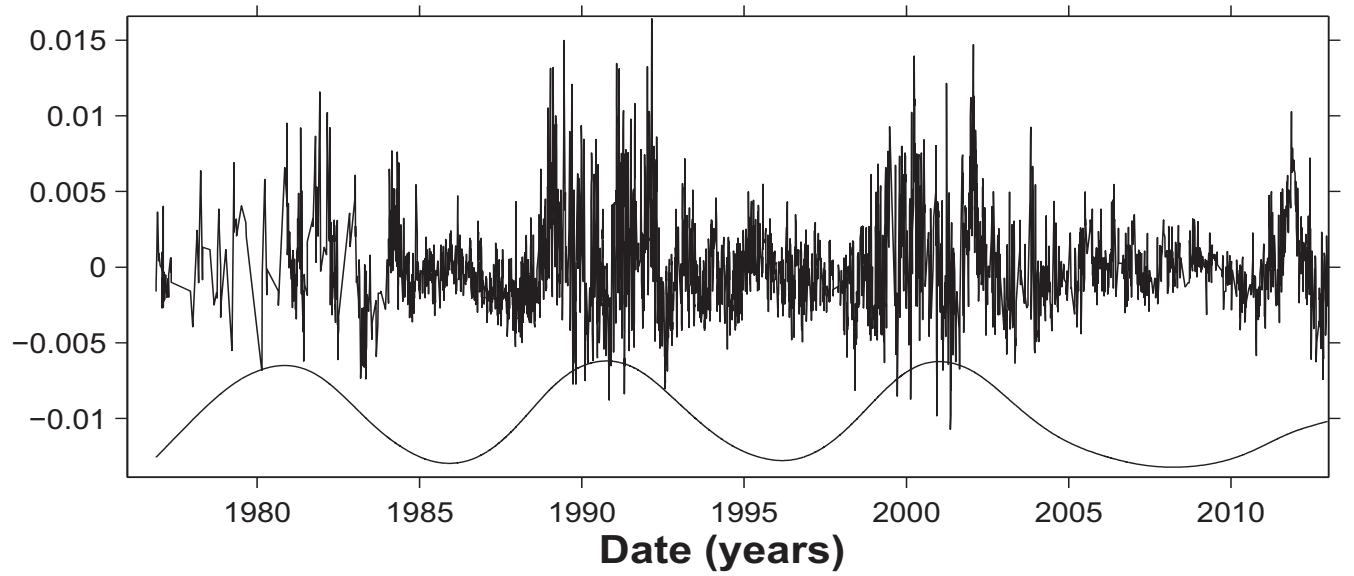


Fig. 6.— Residuals of the **EMDX** data from the adopted smooth fit, which is plotted with an arbitrary offset and scaled down by a factor of 2 relative to the residuals.

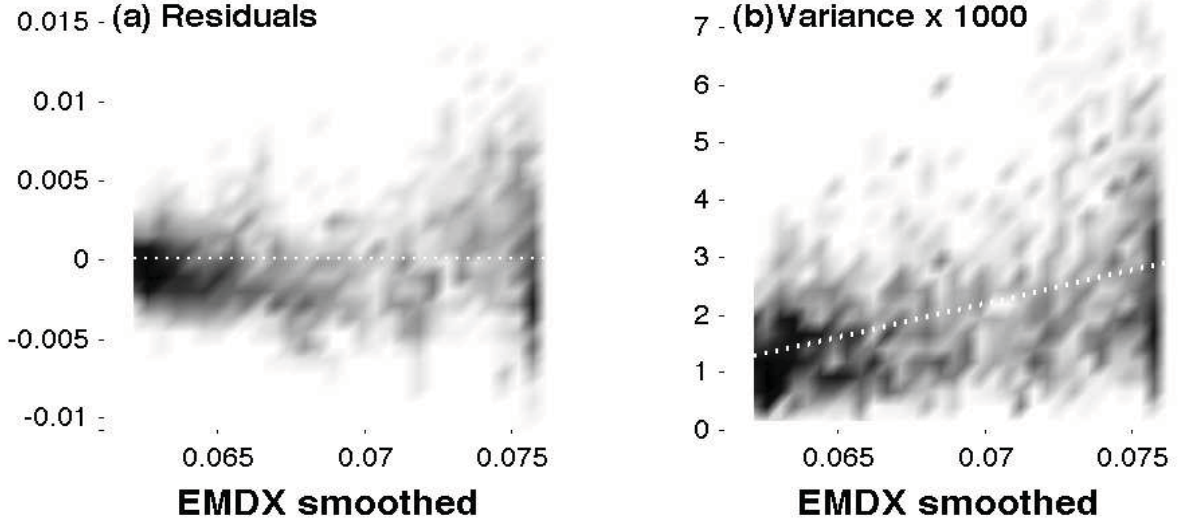


Fig. 7.— **EMDX** residuals (a) and variance of these residuals (b) vs. smoothed **EMDX**. The dotted lines are (a) zero residual and (b) least squares regression.

Figure 7 is another way to visualize the relationship between the random and smoothed **EMDX**. By construction the residuals average to zero, as in (). The increase of the variance with amplitude is explicit in panel (b) and supported by the increase of the range of the residuals with emission in (a).

Table 2 presents summary statistics for the seven variables. Except for the error estimates described in §3 these all computed in straightforward ways directly from the time series data. The first three rows (after one specifying the units for the quantity) contain the mean, range, and standard deviation computed directly from the raw observations with outliers removed. Row four is the standard deviation of the residuals from the adopted smooth fit to the relevant time series. Row five gives the estimated RMS observational errors described in the next section; these should be taken as upper limits for reasons described there. Row six is the relative error obtained by dividing row five by row two. Row 7 is the index α in power-law fits ($P(f) \sim f^\alpha$) to the power spectra, described in §4.

Table 2: Statistics of the K-line Time Series. N.B. these numbers will be updated slightly with data available at the time of publication.

	EMDX	VIORED	K2VK3	DELK1	DELK2	DELWB	K3
Units	Eq W	Δ intensity ratio	intensity ratio	$\Delta\text{\AA}$	$\Delta\text{\AA}$	$\Delta\text{\AA}$	intensity
(1) Mean	0.0929	1.2700	1.5146	0.6304	0.3748	1.5832	0.0687
(2) Range	0.0129	0.2256	0.2648	0.1862	0.0918	0.1031	0.0211
(3) σ	0.0043	0.0444	0.0549	0.0508	0.0204	0.0212	0.0062
(4) Residual σ	0.0018	0.0321	0.0250	0.0229	0.0144	0.0116	0.0024
(5) Error \leq	0.0005	0.0303	0.0133	0.0180	0.0114	0.0108	0.0006
(6) Error \div Range	0.0423	0.1343	0.0501	0.0967	0.1241	0.1052	0.0290
(7) α	-0.303	0.004	-0.175	-0.108	-0.114	-0.009	-0.238

3. Autocorrelation Analysis

An *autocorrelation function* contains information about the memory of the underlying process, be it random or deterministic. This function elucidates connections between the quantity at different times; specifically the autocorrelation function $\rho(\tau)$ characterizes the joint variability at times t and $t + \tau$ averaged over t . (In the next section we will also use the autocorrelation as a handy way to compute power spectra and time-frequency distributions.)

The panels of Figure 8 exhibit the autocorrelation function for **EMDX** (computed using the Edelson and Krolik algorithm (Edelson and Krolik 1988) described in Appendix 3) emphasizing three important time scales. The first panel shows the autocorrelation function (normalized to unity at zero lag) extending to the maximum lag, namely the 13179 day length of the time series, thus emphasizing the time scale of the solar cycle. The bottom two panels show the unnormalized *autocovariance function* (different by only a constant, and indicating actual variances). These plots cover lags in the range of the surface rotation period and the one-day sampling of the raw data, respectively.

The overall behavior of this function is dominated by variability at the frequencies of the solar cycle and the surface rotation. In the top panel much of the scatter about what would otherwise be a smooth curve is due to a combination the stochastic signal (3), the rotational modulation (4), with a minor contribution of the errors (5). The increased scatter for large lags results from fewer data points contributing to the average.

In the bottom right panel of Fig. 8 as one considers smaller and smaller lags the autocorrelation levels out somewhat at two days and one day, and the value at zero lag is notably higher than this level. This offset, also visible in the other panels, reflects the error variance and provides a way to estimate the average observational error. The auto-correlation function at zero lag is the sum of two contributions: the observational error variance and the true variance of the source. At any other lag the errors average to zero as long as they are uncorrelated. These remarks yield a procedure for estimating the size of the average observational error by attributing it to the excess in the zero-lag spike. Assuming the true autocorrelation function is reasonably smooth, the difference between an extrapolation to zero lag and the actual value $\rho(0)$ yields the variance corresponding to the

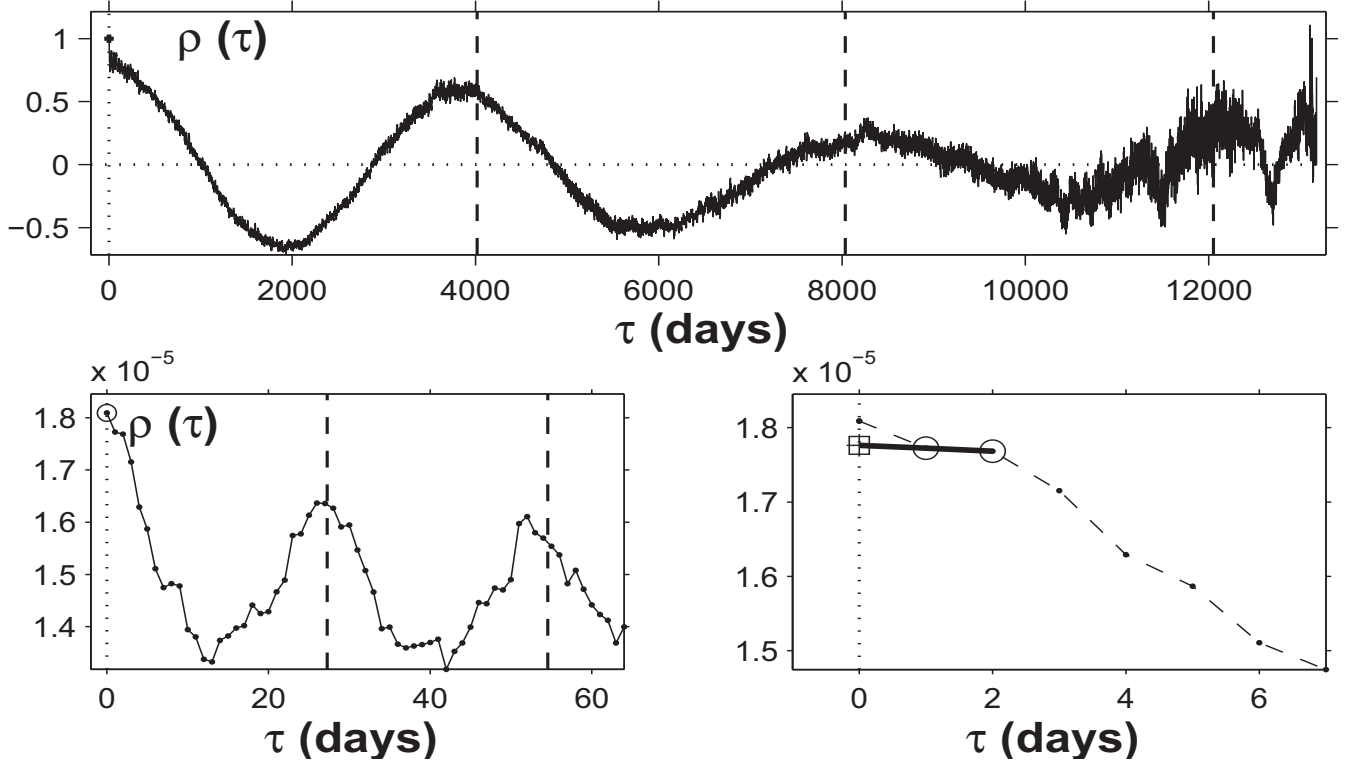


Fig. 8.— Edelson and Krolik autocorrelation function of **EMDX** time series, with zero lag indicated by vertical dotted lines. Top: Autocorrelation (normalized at zero lag), over the full range of lags. Bottom: unnormalized autocovariance at lags up to 64 days. Lags of multiples of 11 years (top) and the Carrington period of 27.2753 days (bottom left) are shown as dashed vertical lines. The bottom right panel illustrates determining the error variance by subtracting the projected from the actual zero-lag ACF.

observational errors:

$$\sigma_{err}^2 = \rho(0) - \lim_{\tau \rightarrow 0} \rho(\tau) . \quad (1)$$

In the bottom-right panel of Figure 8 a linear fit to the autocorrelation at the first two positive lags (shown as circles) was extrapolated to the point contained in a square. Clearly an even smaller estimate would result from an extrapolation from three or more points. For this and the other parameters the difference between the estimated and extrapolated zero-lag values is taken to be the error variance and reported in Table 2.

Actually these error estimates should be regarded as upper limits. Any solar variability

confined to time-scales shorter than the sampling interval would make a contribution to the zero-lag spike that would be lost in our extrapolation procedure. Although known turbulence and oscillations likely contribute in this way, absent more quantitative information on how smooth the true autocorrelation function is on the diurnal time scale, the errors listed in Table 2 are probably conservative upper limits.

4. Power Spectrum Analysis

Rotation can produce a periodic modulation of any solar time series. This signal might be expected to be relatively weak in full-disk observations of chromospheric lines, as discussed further in §5. Nevertheless one goal of this work is to detect and characterize any signatures of rotational modulation in the K-line time series; *c.f.* (Bertello et al. 2012). This section demonstrates the rather strong rotational signal present in these data and already remarked upon in §2 in connection with Figure 5. Rotation yields peaks in the power spectrum at the solar rotation frequency and its harmonics. Indeed even the more subtle effects of differential rotation can be studied in considerable detail, as shown in the following section.

Figure 9 shows two power spectra for the **EMDX** time series, both computed using the Lomb-Scargle periodogram (Gottlieb et al. 1975; Lomb 1976; Scargle 1982; Scargle 1989). Power spectra computed from the Edelson and Krolik auto-correlation function, as mentioned in §3 and detailed in Appendix 2, are essentially identical to those shown here. The comparison is between the spectrum of the raw data (top) and that of the residuals from the smooth fit (bottom). The modulation at the rotation frequency and its harmonics is largely buried in three processes: the noise inherent in unsmoothed power spectra, the longer time scale variability, and the noise in the raw data. It is slightly more prominent in the power spectrum of the residual data in the bottom panel. It is a textbook result that raw power spectra are noisy and require smoothing in order to reveal much of their information content; *cf.* (Scargle 1982). We do not pursue this avenue here, since an even more fruitful approach for the current application is detailed in the following section.

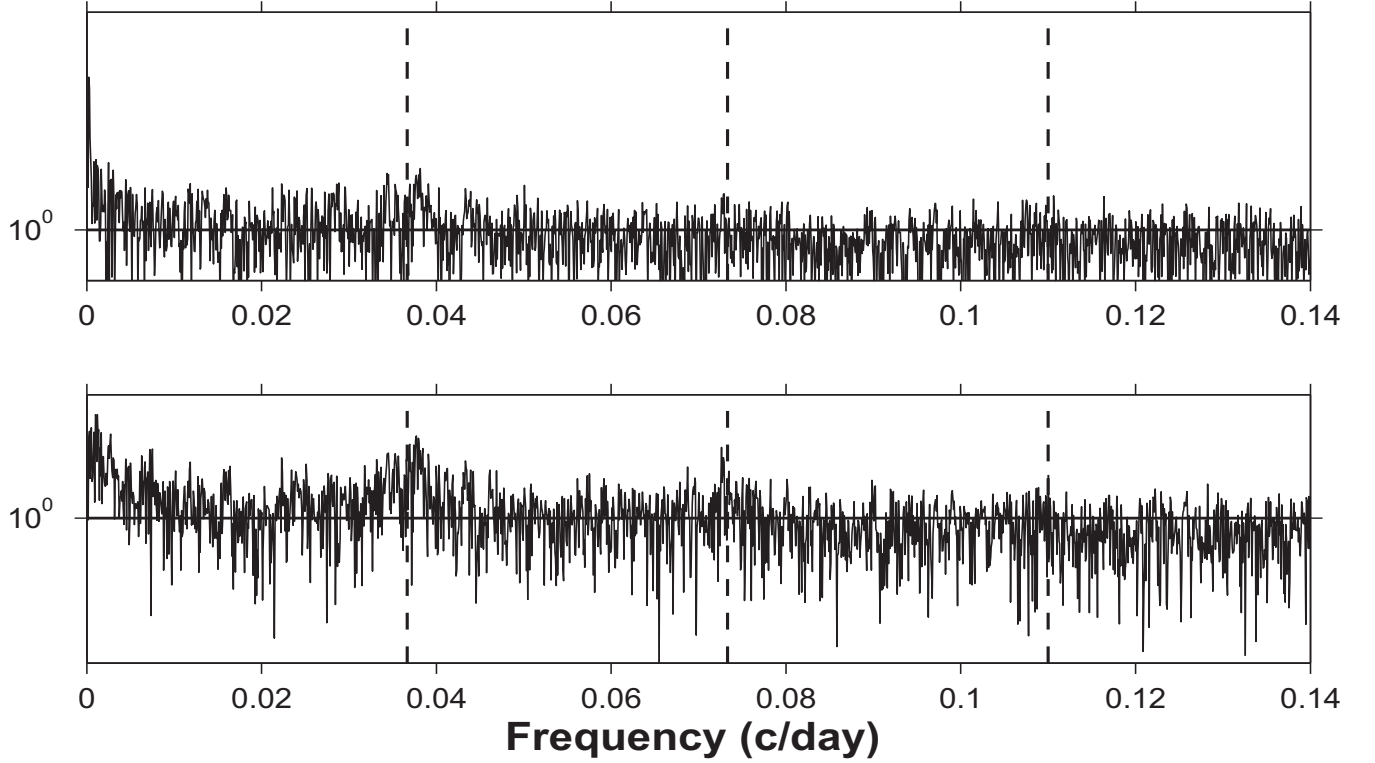


Fig. 9.— Lomb-Scargle power spectra of: (top) the raw **EMDX** data (top) and (bottom) the residuals from smooth fit as shown in Figure 6. Vertical dashed lines denote the nominal frequency (corresponding to the Carrington period of 27.2753 days) and harmonics.

5. Time-Frequency Distributions

Rotation induces a harmonic variation of any signal from a localized region of the Sun’s surface and observed from a fixed direction.¹ In the case of spots and active regions differential rotation modulates the observed period as they experience the latitude drift tied to the solar cycle. The summation of sources at various solar longitudes and latitudes inherent to full disk observations smears out the power spectrum and dilutes the signature of differential rotation. Nevertheless considerable information about the Sun’s rotation is contained in the full-disk K-line time series, as we shall now see.

The *time-frequency distribution* is designed specifically to explore this sort of evolving harmonic structure. In a nutshell this signal processing tool displays the time evolution of the power spectrum. The excellent treatise (Flandrin 1999) explains what can be learned with the basic tool and a number of its variants. Here we use the simplest approach, namely computing power spectra of the data within an ordered sequence of time windows framing sub-intervals of the full observation span. Accordingly this tool is also called a *dynamic* or *sliding window* power spectrum. The output is a three dimensional plot – power (z-axis) as a function of time and frequency (x- and y-axes) – that we here render as 2D grayscale plots.

A slice of this plot parallel to the frequency axis contains the power spectrum (power vs. frequency) at a specific time. A slice parallel to the time axis depicts the time dependence of power at a specific frequency. The same mathematics leading to the Heisenberg uncertainty principle dictates that the time and frequency resolutions are not independent and cannot both be made small simultaneously. Good frequency resolution can be had only with relatively large time windows, and good time resolution requires short windows.² Most implementations of the time-frequency distribution allow one to mediate this unavoidable

¹Observed from the Earth the rotational modulation of a single such region is approximately a truncated sinusoid, with Fourier components at the frequency corresponding to relevant synodic period, plus harmonics.

²Simply decreasing the time increments by which the window is moved does not increase the time resolution. It is the size of the window that fixes the smoothing in the time-domain.

resolution trade-off, for example by adjusting the size of the window. Some details of the computation of time-frequency distributions are given in Appendix 2.

Figure 10 shows time-frequency plots for the emission index. Because the data are not evenly spaced, all of the time-frequency distributions shown here were computed using the technique described in Appendix 2. This approach does not require any interpolation, thus avoiding the corresponding information loss. The cross symbols near the top right corners indicate the time and frequency resolution. The length of the arms of the cross indicate the width of the sliding window and the corresponding fundamental frequency. The vertical lines mark the frequency range corresponding to the rotation rate as a function of solar latitude, using Equation (3) of (Brown 1989):

$$\frac{\Omega_p}{2\pi} = 452 - 49\mu^2 - 84\mu^4 \quad \text{nHz} , \quad (2)$$

labeled with the corresponding latitudes (in the 0 to 60 degree range normally inhabited by spots).

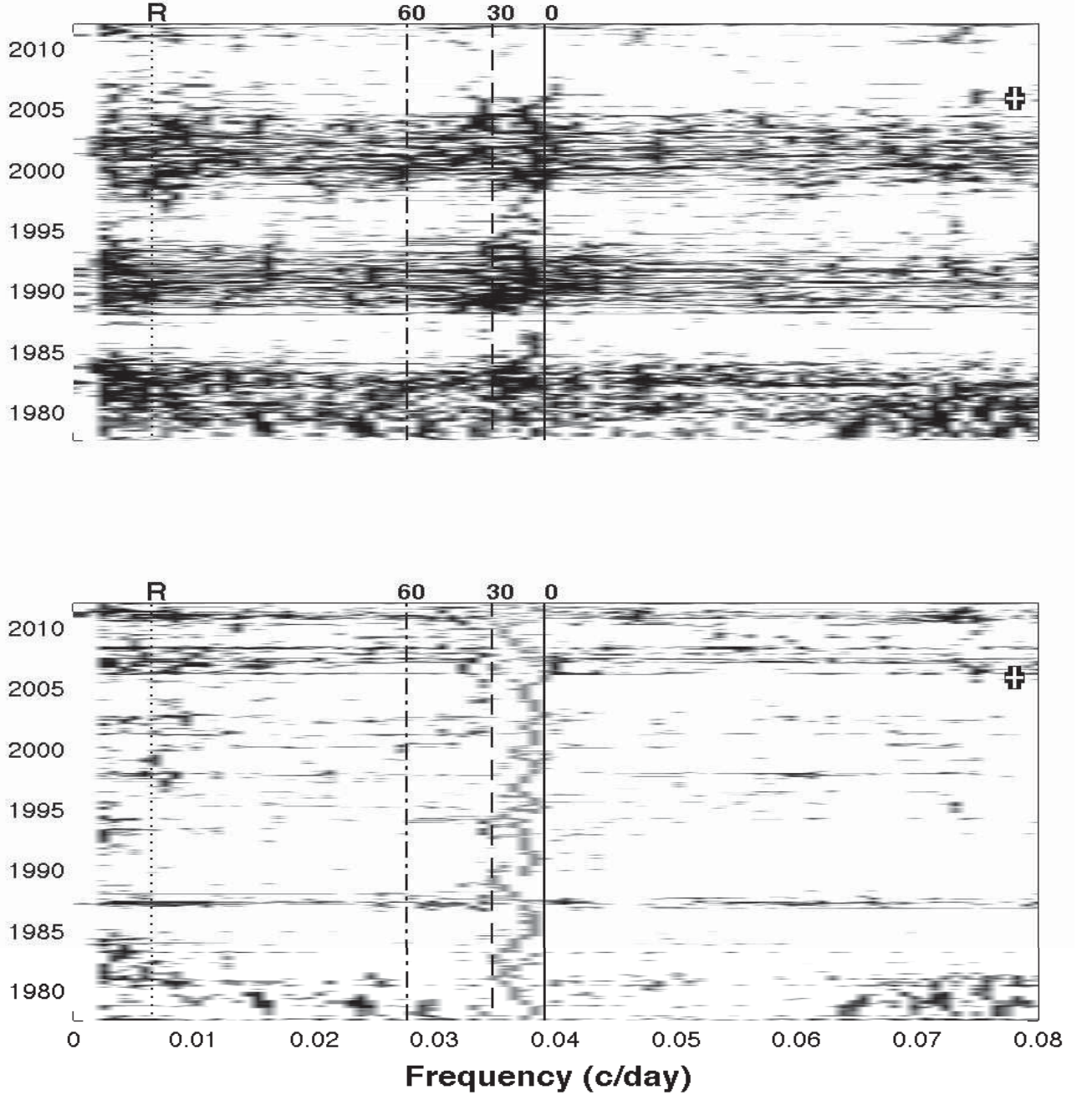


Fig. 10.— Time-Frequency distribution of **EMDX** residuals. Time and frequency resolutions are indicated by the cross at the top right. The plot in the bottom panel is renormalized to make the power between the frequencies corresponding to 0 and 30 degrees constant, bringing out behavior during solar minimum that is otherwise lost. Solid, dashed, and dot-dashed vertical lines denote latitudes 0, 30, and 60 deg, respectively, based on Equation (2). The dotted line (labeled R) at .0082 c/d roughly corresponds to quasi-periodicities discussed in §4. Power below .00176 c/day is divided by 10 to improve the display.

In the top panel of Figure 10 the rotational features almost disappear during solar minimum and are generally strongest in the middle cycle 22. It is instructive to correct for these effects by renormalizing each time slice of the distribution, as in the bottom panel. Note the interesting behavior of the rotational signal in this simply rescaled version of the top panel: a component moves back and forth between approximately 0 degrees and 30 degrees latitude and is present essentially all of the time, not just during solar maximum as one might have concluded from the top panel. A similar conclusion was reached in (Bertello et al. 2012).

The signal processing literature contains descriptions of many other ways to estimate time-frequency distributions (Flandrin 1999). One of the most recent ones, called *synchrosqueezing*, represents the time series as “the superposition of a (reasonably) small number of components, well separated in the time-frequency plane, each of which can be viewed as approximately harmonic locally, with slowly varying amplitudes” (Daubechies et al. 2009). Figure 11 shows the result of this analysis for **EMDX** and **K3** interpolated to even spacing, using the MatLab tools in (Brevdo and Wu 2011) and described in (Brevdo et al. 2011; Thakur et al. 2012). The gray scale represents the power spectrum of the variables, as a function of time and frequency. There are broad similarities to the distributions in the previous figure, and the differences in detail can be understood in terms of the constraints imposed by the synchrosqueezing method on the time-frequency atoms, as well as the fact that interpolation to even spacing was necessary. The more fully non-parametric nature of the sliding-window Fourier spectrum yields a different representation of the underlying time-frequency structure, and with a different tradeoff between time and frequency resolution. The synchrosqueezing algorithm renders the differential rotation features as more discrete and less continuous than does the sliding-window approach; the reverse is true of the quasi-periodic signals periods in the vicinity of 0.002-0.015 cycles per day.

6. Cross- Analysis

Each of the seven K-line parameters probes a different aspect of the chromosphere. Therefore relations between the corresponding time series can elucidate physical processes

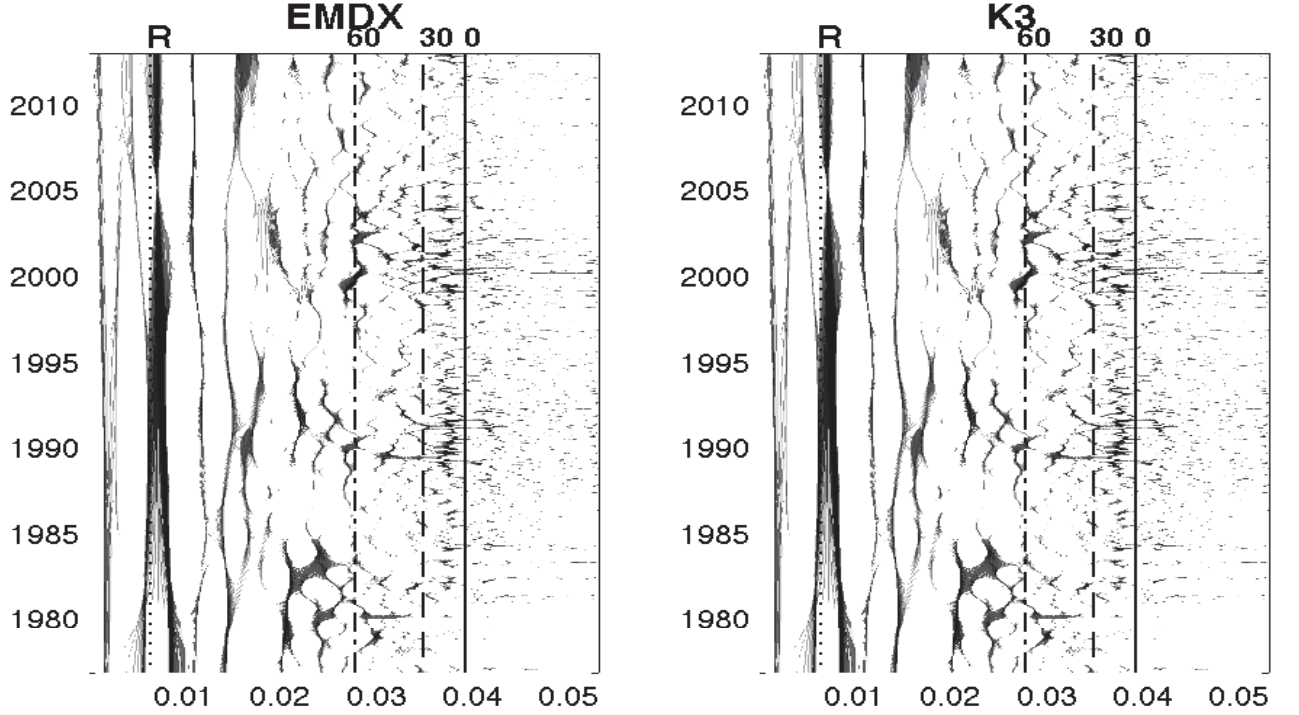


Fig. 11.— **EMDX** and **K3** time-frequency distributions computed with the synchrosqueezing algorithm. The number-of-voices parameter = 50. The vertical lines represent the same nominal frequencies as in Figure 10.

underlying chromospheric activity. For example variability in two parameters that is correlated, anti-correlated, or correlated with time-lags can shed light on underlying dynamical processes. Of course causality cannot be proven in this way, but relationships consistent with physical models can be elucidated.

Figure 12 depicts two types of cross-analytic relationships for all pairs of parameters. The 21 scatter plots above the diagonal describe inter-dependence among the parameters. Below and on the diagonal are cross- and auto-correlation functions, respectively; all were computed with the Edelson and Krolik algorithm (*cf.* §3 and Appendix 3). The autocorrelations are symmetric, but both sides have been plotted for scale consistency with the cross-correlations.

These types of displays are complementary ways of relating two variables. Independence

is a stronger statistical condition on two variables than their being uncorrelated. The former implies the latter, but not *vice versa*. Hence to the extent that scatter plots elucidate dependence they are more powerful statistically. However they depict only simultaneous relationships, whereas cross-correlation functions elucidate how the variables at two different times are related.

In computing the quantities displayed in this figure all seven parameters (with outliers removed) were standardized to zero mean and unit variance; the data in Table 2 can be used to restore the original values if desired.

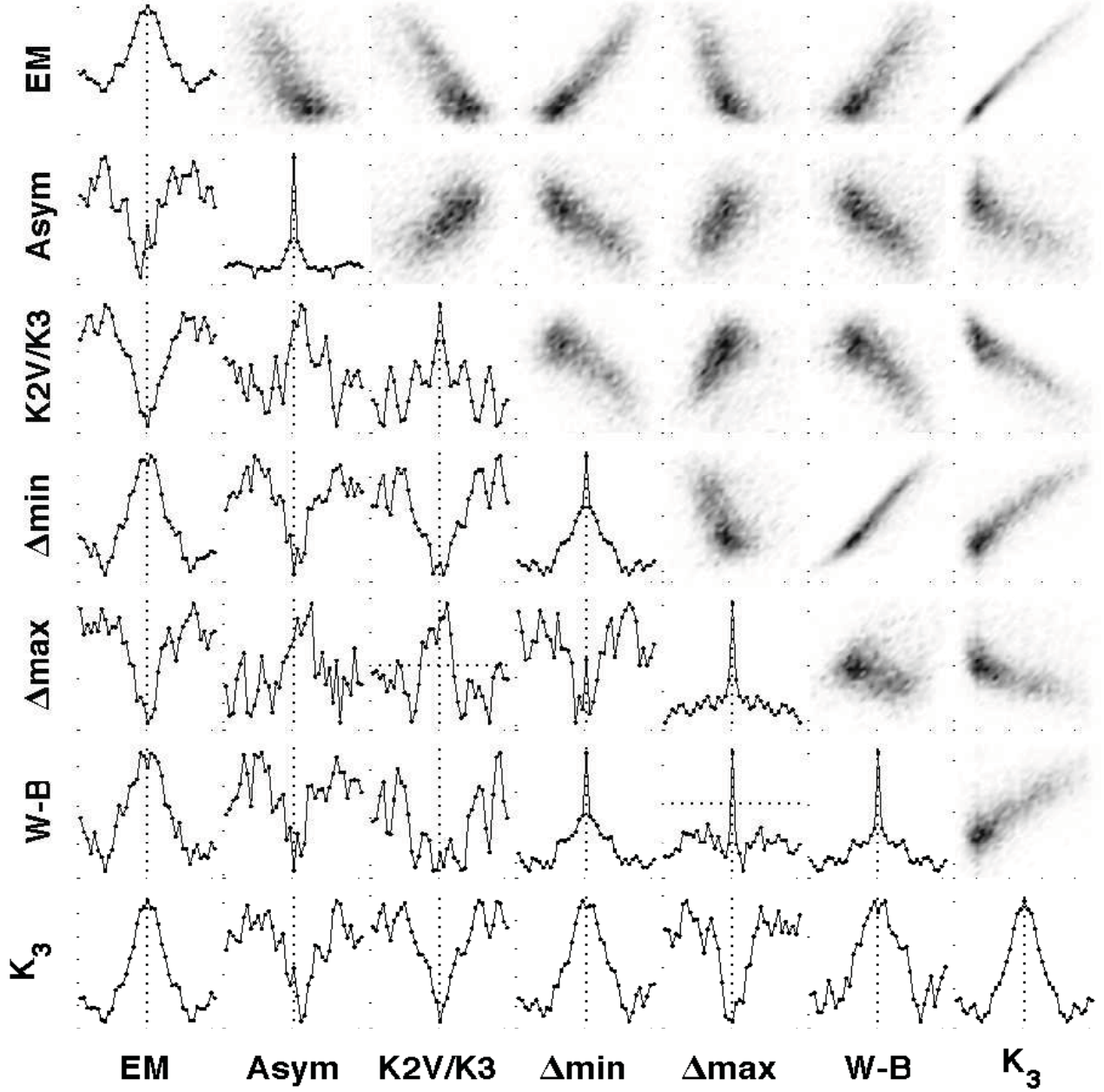


Fig. 12.— Diagonal: Auto-correlations. Above Diagonal: Scatter plots in the form of density vs. the simultaneous values of the two parameters, rendered as grayscale plots. Below Diagonal: Cross correlations. All correlations are plotted for $|\tau| \leq 20$ days.

7. Wilson-Bappu Effect

There is evidence for a Wilson-Bappu effect in these data, in the sense that the intensity parameters correlate with the width of the K-line. Figure 13 contains scatter plots for the four intensity parameters vs. the Wilson-Bappu parameter, computed as regular 64 by 64 two-dimensional histograms and portrayed as greyscale plots. The left-hand column contains scatter plots for the raw data (with outliers removed). These correlations are presumably due to chromospheric processes tied to the variations in physical conditions over the solar cycle. The right-hand column shows the corresponding residuals from the smooth fits described in §2, which are essentially uncorrelated.

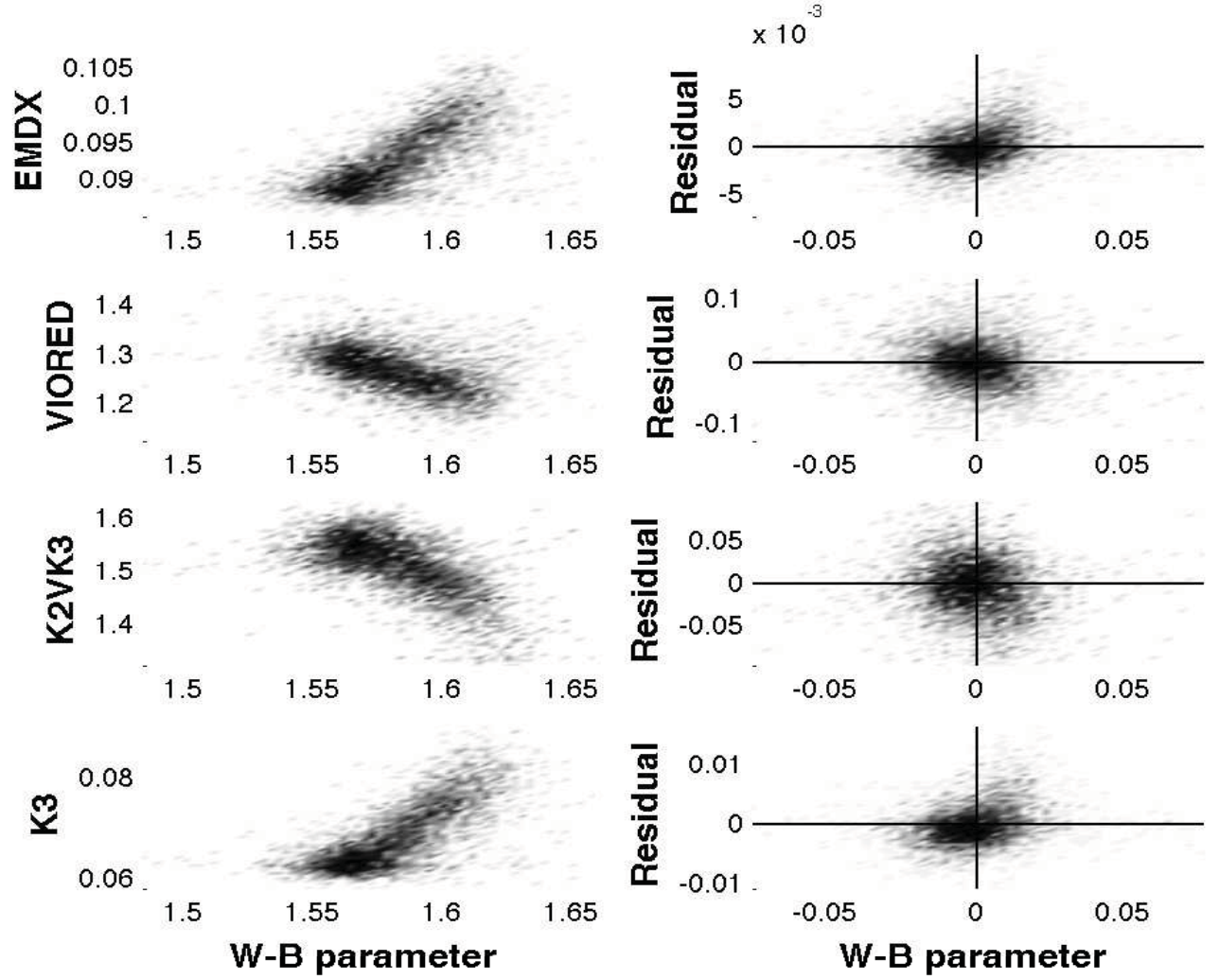


Fig. 13.— Intensity variables vs. Wilson Bappu parameter. Left-hand column: Raw Measurements (with outliers removed). Right-hand column: residuals from the spline fits.

8. Conclusions

The goal of this exploratory study of the data collected in the K-line monitoring program is to follow the clues provided by various time series analysis methods in the time-, lag-, frequency- and time-scale domains.

Affecting the interpretation of any such results is their significance in light of random observational errors and systematic errors. Even though point-by-point errors are not available, we have placed reasonable upper limits on the average error variance in §3. It is difficult to construct and display errors on 2D and especially 3D relations such as time-frequency distributions. We rely on comparison of these functions for various parameters to indicate the rough importance of observational errors. For example the time-frequency distributions for **EMDX** and **K3** for the most part show the same behavior and therefore indicate that such errors do not materially affect the results. The other parameters show significantly different behavior, but comparison of similar ones (such as the three wavelength differences) gives a similar indication of the significance of the time-frequency structures.

The conclusion that the solar cycle is complex, as discussed in §2, is not new but perhaps a refinement of concepts of the sunspot cycle and other heliospheric phenomena discussed by a number of authors. Apparently the K-line features are particularly sensitive to the changing physical conditions during the solar cycle. A carefully chosen degree of smoothing of the time series is essential to elucidating this structure. Previous discussions have at most centered on a relatively simple double peak structure in sunspot and other heliospheric indices at the time of solar maximum.

For example, in yearly averages of the number of intense geomagnetic storms (Gonzalez et al.1990) describes time series behavior that generally follows the solar cycle, but with a double peak structure: “ ... one at the late ascending phase of the cycle and another at the early descending phase,” with hints of even more complex three-peaked structure for cycle 20.

In (Hathaway 2010) Figures 16 and 38 show distinct double peak structures in the smoothed International Sunspot Number, and Figure 42 shows the same for the Polar Magnetic Field Strength as measured at the Wilcox Solar Observatory. One presumes that

the weakness of apparent double peaks in sunspot number averaged over cycles 1 to 22 depicted in this paper could be because of slight variations of the locations of the peaks as well as the degree of smoothing applied to the individual curves.

Figure 9 in (Domingo et al. 2009) depicts a double peak and more complex structure in the time series for sunspot numbers, 10.7 cm. radio flux, a Ca-K 1 \AA full-disk index from Kitt Peak, and a Mg II index. The reference (Khramova et al. 2002) gives an overview of structure in sunspot variability on various time scales, referring to the kind of structure noted here as “quasi-biennial oscillations.” The complex time series structures shown in the bottom panel of our Figure 4 perhaps correspond to multiple toroidal field surges as discussed by (Georgieva 2011).

A somewhat related issue is the structure noted in the time frequency distributions, possibly connected to the solar cycle but at frequencies lower than those due to surface rotation. There is discussion in the literature of such frequencies, often in the context of an early claim of a 154 day periodicity in solar gamma-ray flares (Rieger et al. 1984), which was followed by attempts to find similar periods in other phenomena. (Sturrock 1996) discusses an idea in which a more complex structure consisting of multiples of a fundamental period of approximately 25.5 days underlies the Rieger periodicity; see also (Bai and Sturrock 1993). (Hill et al. 2009) discusses a period of 151 days in solar cosmic ray fluxes. (Joshi and Joshi 2005) finds a 123-day period in soft x-ray flux from the sun, and (Lou et al. 2003) finds a very similar period (and others) in solar coronal mass ejections. The relevance of similar periodicities occurring in other stars (Massi et al. 2005) is unclear.

Our K-line data as analyzed in the time-frequency distributions in §5 suggest the presence of some quasi-periodic behavior on similar time scales. We found that lines in a sine wave of suitably chosen period and phase matches many of the peaks in the partially smoothed **EMDX** time series. A period of 122.4 days was obtained in a rough peak-fitting procedure. However, keeping in mind the degrees of freedom in the sinusoid, the uncertainty in locating and timing the peaks in the data, and the matching of only some peaks, this result does not prove the existence of a pure harmonic signal. Rather as indicated in the time-frequency distributions there appears to be quasiperiodic behavior in this frequency range.

Another result of our analysis is the characterization of differential surface rotation, using the time-frequency tool (§5) for the main intensity variables, with displays for all of the variables in Appendix 1. For discussions of differential rotation estimated from time series from Kepler and CoRoT see (Frasca et al. 2011) and (Silva-Valio 1990) respectively.

Other conclusions include:

- The solar cycle variability [component (1) in Table 1] of the K-line intensity consists of a broad oscillation consonant with the well known 11-year and 22-year cycles. The current cycle – in relation to the previous three – has started late and may also turn out to be weak.
- In addition there are quasi-periodic oscillations that do not have constant periods or amplitudes, but irregularly populate the time-frequency domain in the neighborhood of periods of 100 days. It is not clear if these are modulations of the solar cycle or an independent physical process.
- The random variations [”Flicker noise”; component (2) in Table 1] in the intensity parameters have a power spectrum describable as $\frac{1}{f}$ noise, meaning $P(\nu) \approx \nu^\alpha$, with $\alpha \approx -0.3$. The other parameters exhibit similar $\frac{1}{f}$ -like behavior, except that the Wilson-Bapu parameter appears to be less correlated, white noise,
- Components (1) and (2) are independent of each other, except that the variance of (2) is correlated with the (1) [Figures 6 and 7].
- A signature of differential surface rotation is captured by time-frequency analysis of the parameters. While this behavior roughly mirrors the general character the butterfly diagram for sunspots, in detail it is distinctly different and is present at all times, not just at solar maximum.

These conclusions refer mainly to the parameters **EMDX** and **K3**, but to some degree apply also to others of the measured parameters.

Acknowledgements: We are grateful to Luca Bertello for helpful comments on the MS. This work was supported by Joe Bredekamp and the NASA Applied Information Systems

Research Program. We are grateful to Alexander Kosovichev and Leif Svalgaard for helpful discussions.

9. Appendix 1: Power Spectra and Time-Frequency Distributions

This appendix presents power spectrum analysis of all of the measured variables. In each of the seven figures the upper-left panel shows the linear Edelson and Krolik-based power spectrum of the residual time series. The vertical dashed lines indicate frequencies corresponding to the 27.2753-day nominal (Carrington) surface rotation period and integral multiples of it.

The lower-left panel shows the time-frequency distribution for the time series, obtained by computing the power spectrum in a time-window that is slid along the data. These were normalized much as in the bottom panel of Fig. 10, but over the broader frequency range 0.02-0.04 cycles/day. The power spectra shown in the top two panels were obtained by averaging this time-frequency distribution over the time coordinate. Accordingly they are considerably smoother than would be obtained directly from the time series. In both of these left-hand panels the full frequency range extending to 1 cycle/day is not shown because most of the interesting behavior visible in a linear plot is below 0.32 cycle/day.

The upper-right panel shows a log-log plot of the full frequency range of the same spectrum. A $\frac{1}{f}$ power spectrum corresponds to a straight line in such a representation. The dashed line is a least-squares fit to the points. The bottom-right panel shows the temporal variation of the slope of the above power-law fit to the power-spectrum – that is the value of α in a representation of the form

$$P(f) = P_0 f^\alpha . \tag{3}$$

Note: we adhere to the convention that a process that even approximately satisfies this equation with any value of α (almost always negative) is called “ $\frac{1}{f}$ noise”.

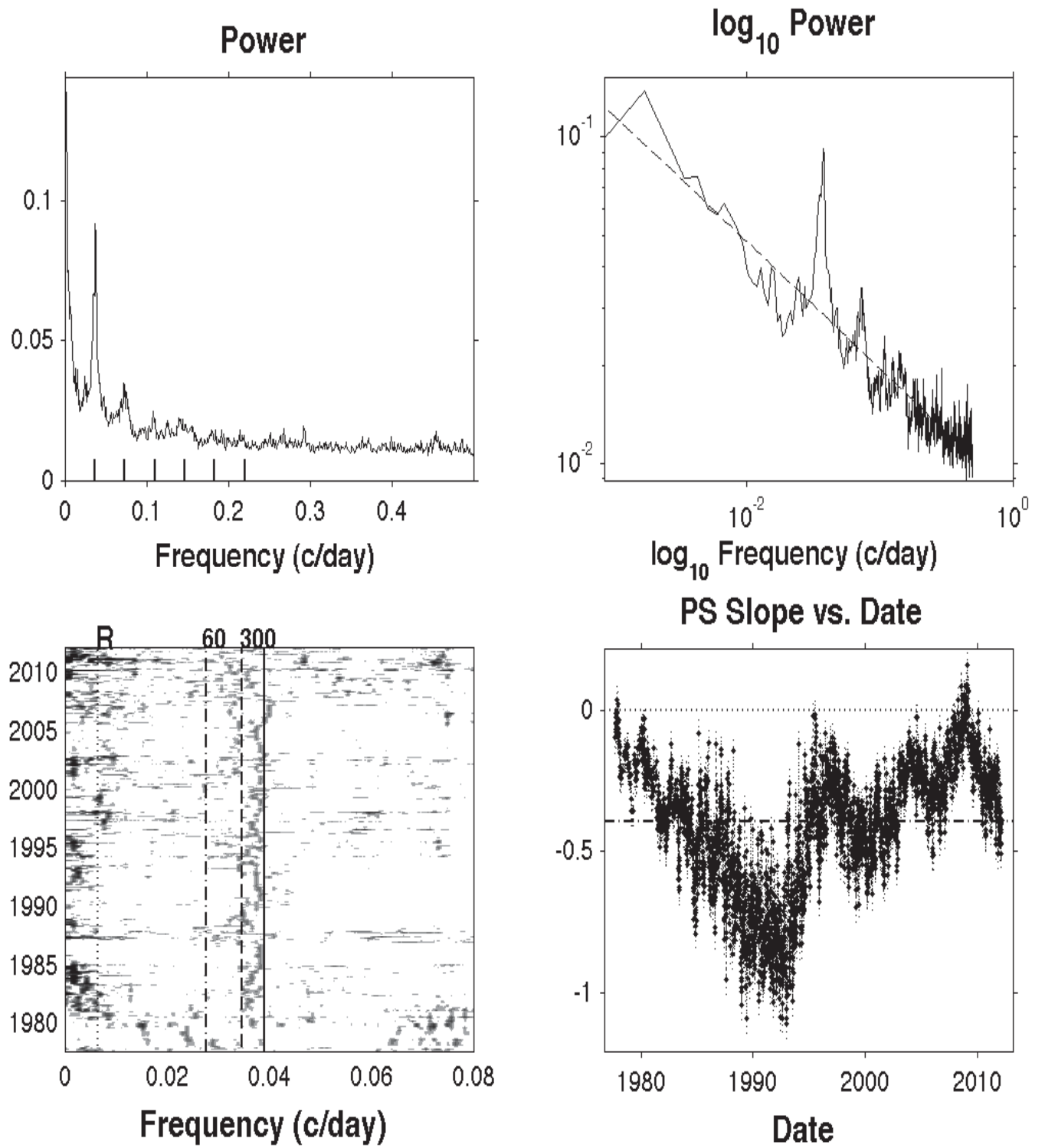


Fig. 14.— Power spectra of **EMDX**.

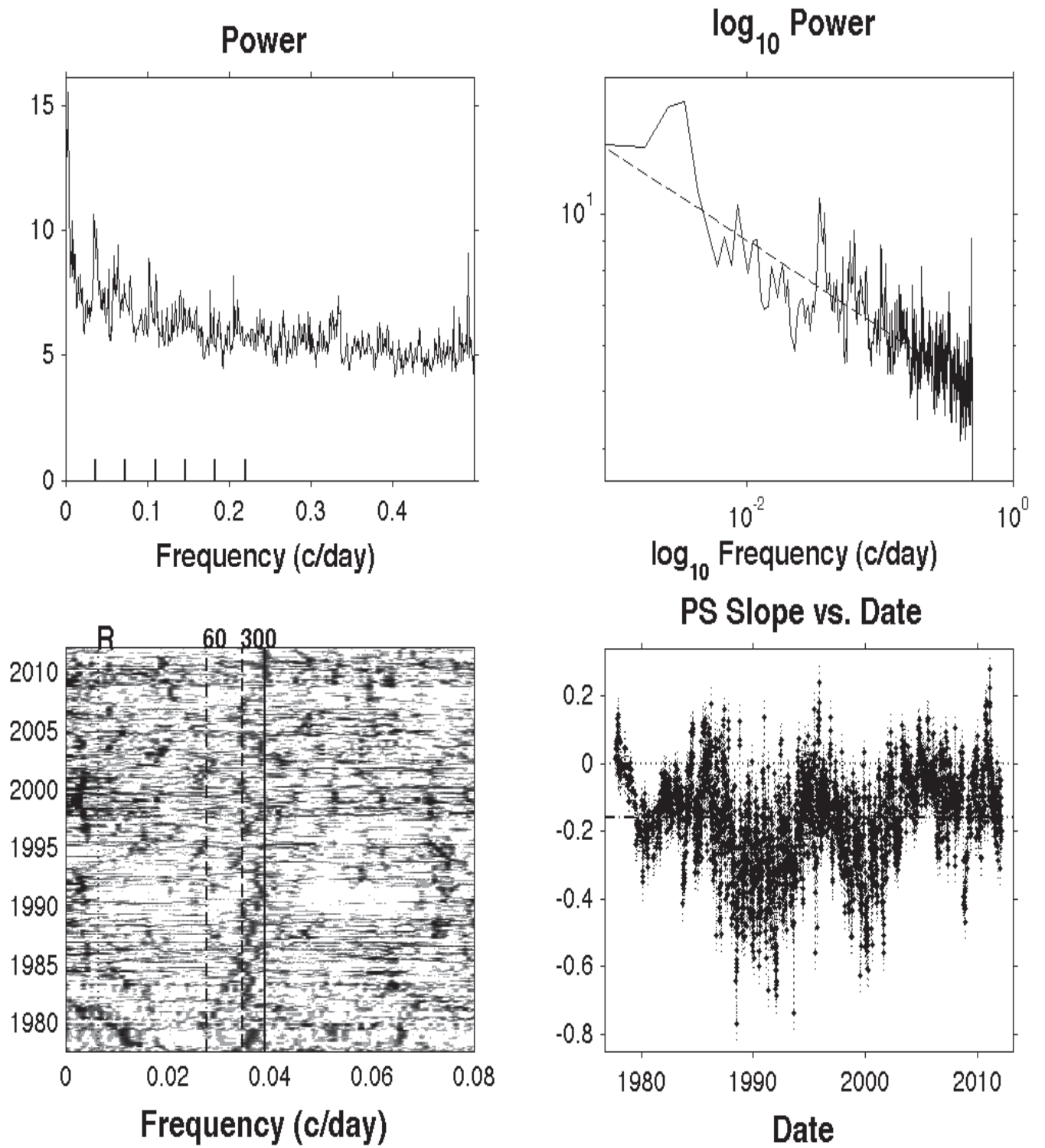


Fig. 15.— Power spectra of **VIARED**.

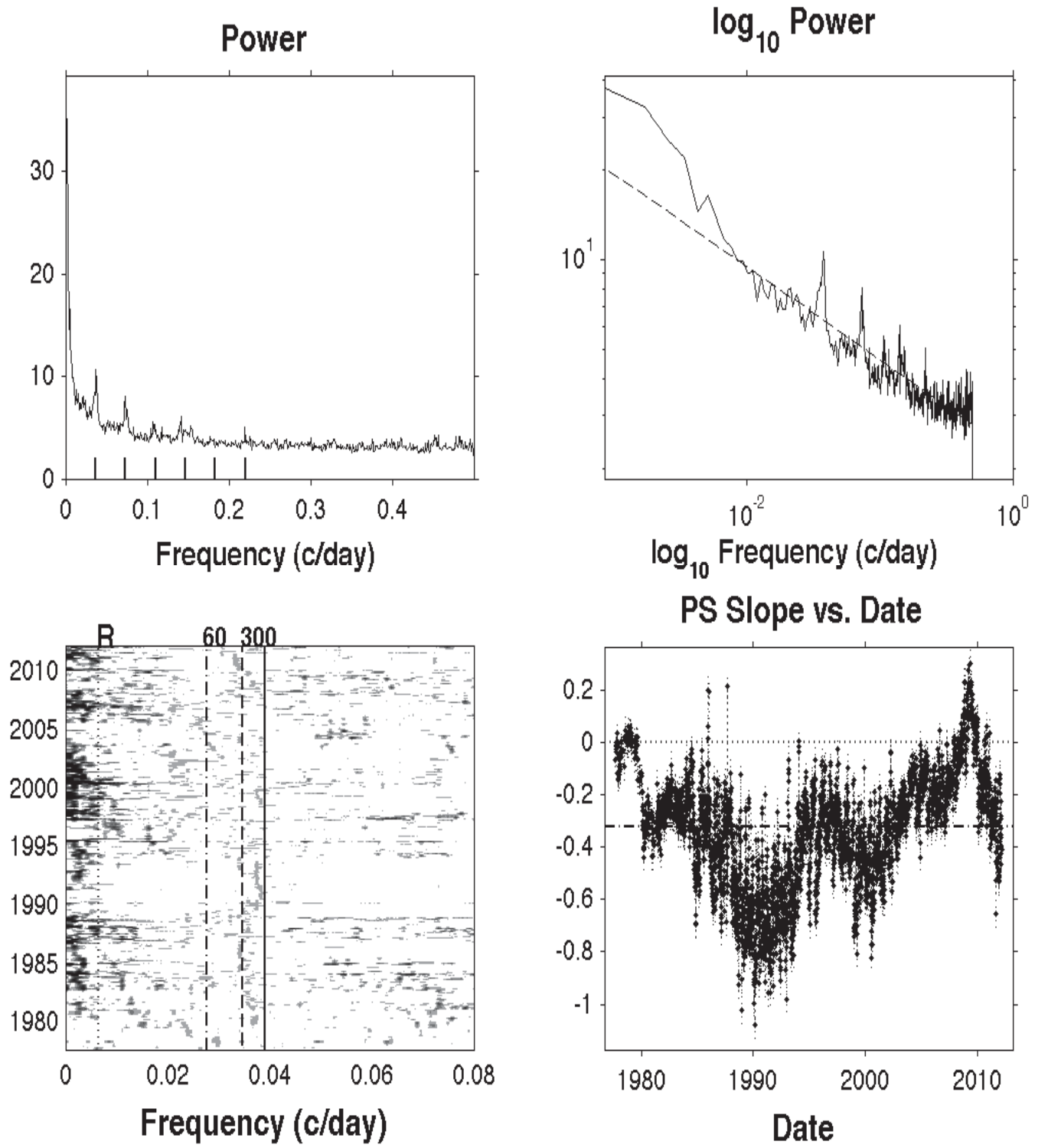


Fig. 16.— Power spectra of **K2VK3**.

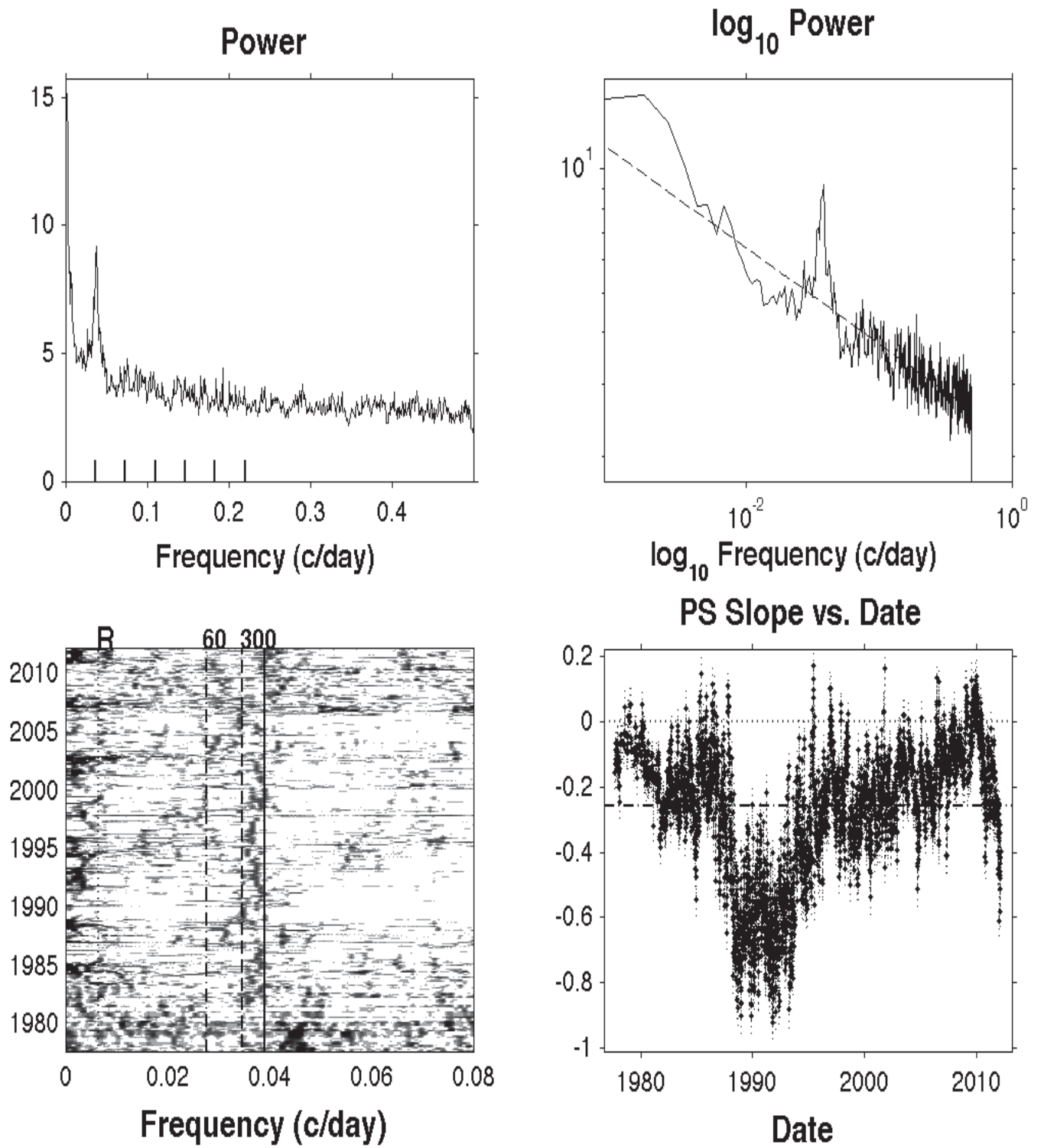


Fig. 17.— Power spectra of **DELK1**.

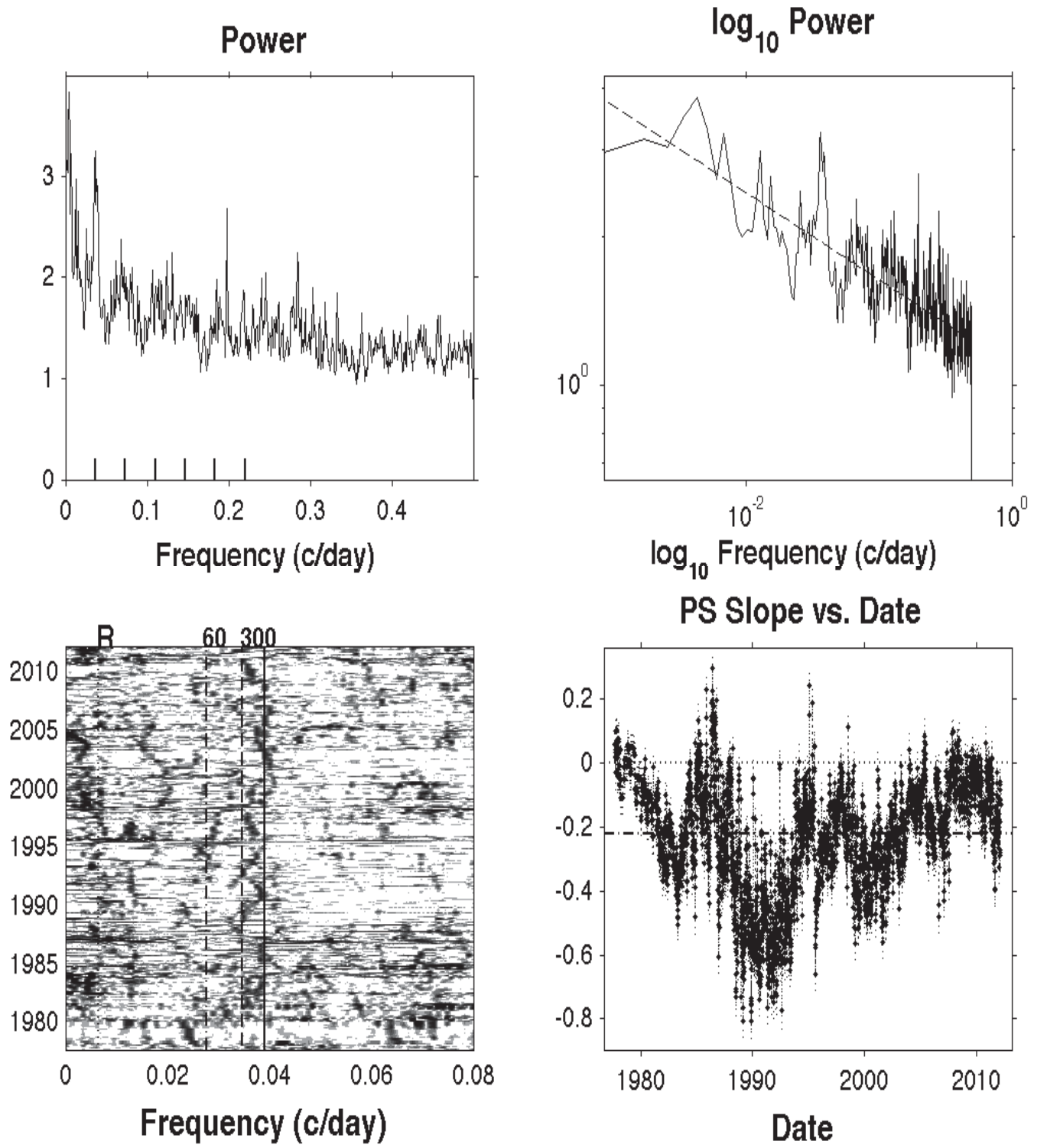


Fig. 18.— Power spectra of **DELK2**.

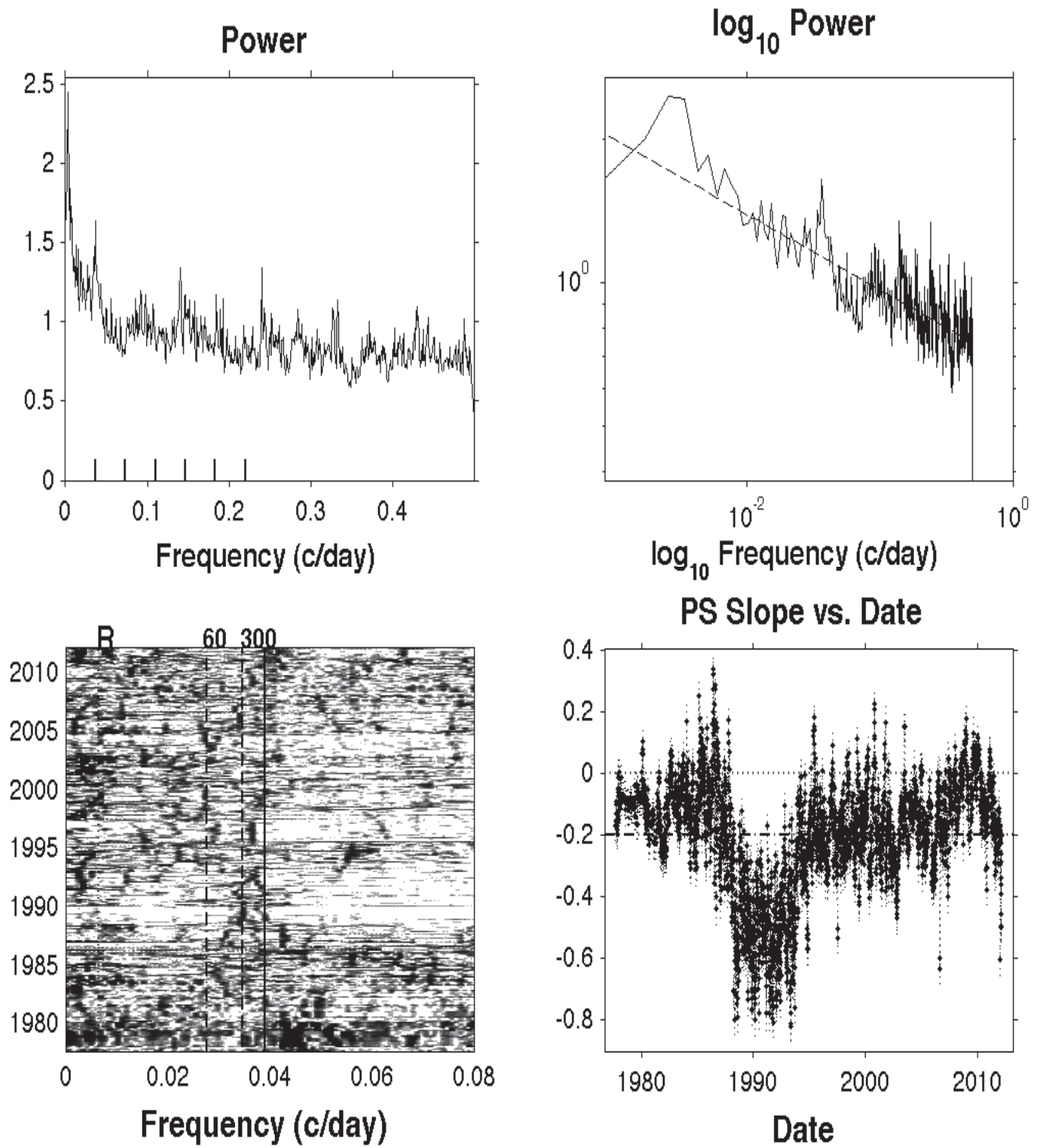


Fig. 19.— Power spectra of **DELWB**.

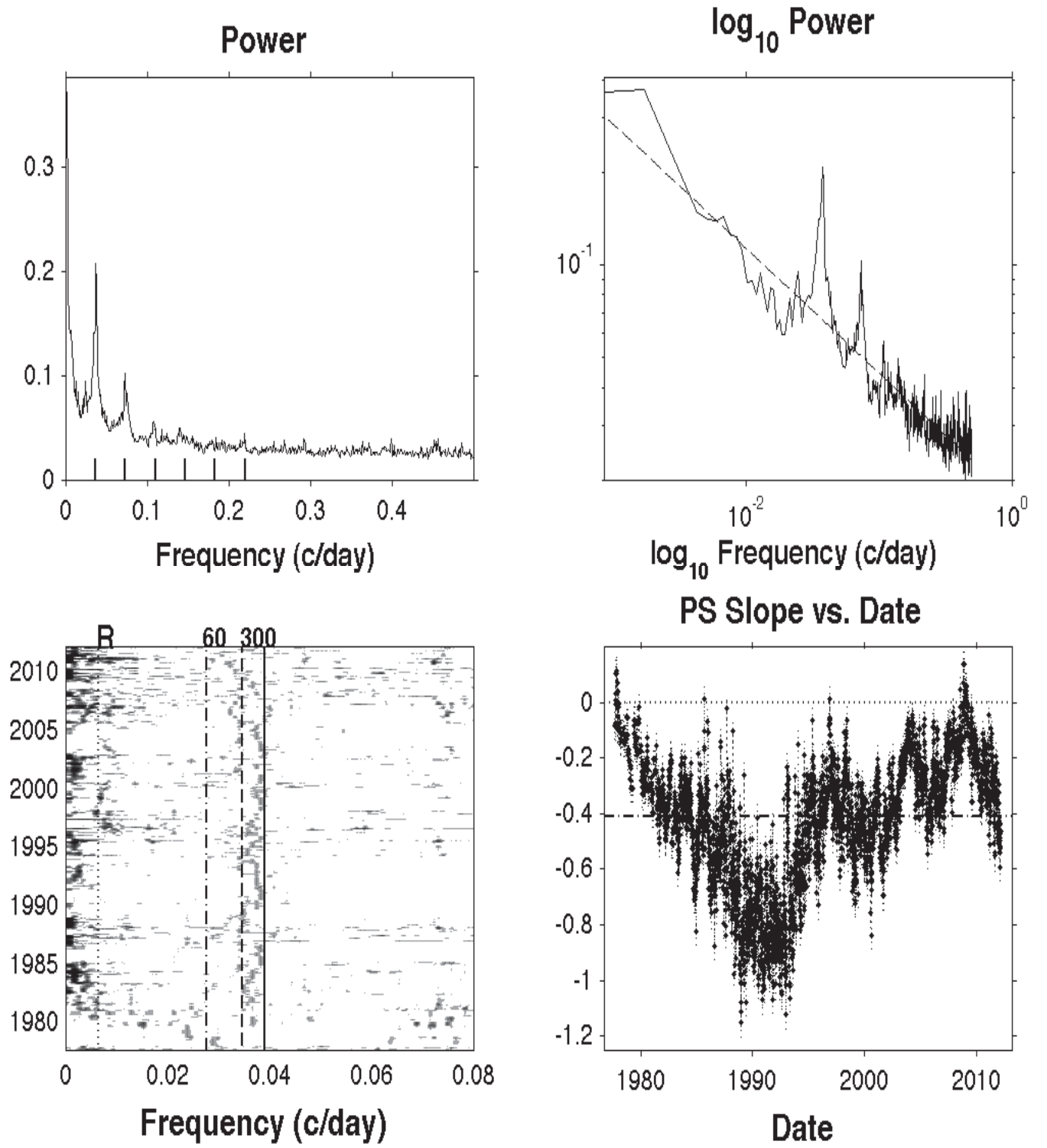


Fig. 20.— Power spectra of **K3**.

10. Appendix 2: Notes on the Computations

These time series represent a special case of irregular sampling, namely evenly spaced (at 1-day intervals) but with some missing observations. The degree of departure from even sampling is depicted in Figure 21.

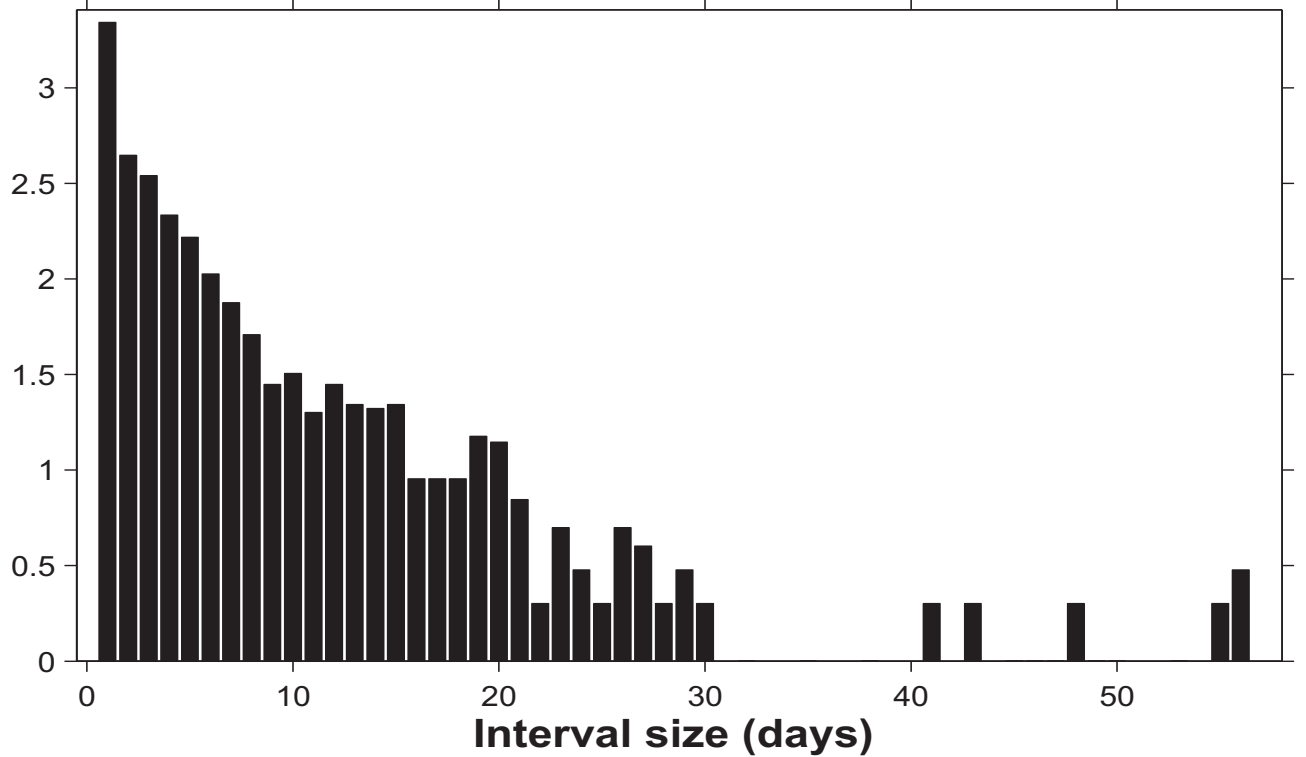


Fig. 21.— Distribution of data gaps: Logarithm (base 10) of the number of cases vs. interval size. For clarity this histogram omits nine large gaps (62 70 71 82 83 123 130 216 174 days) from the early years of the program, before 1982.5. The spike of one-day intervals (2198 out of 3893, or 56.5 per cent) refers to the nominal sampling; anything larger is a gap.

Because of the non-trivial number of gaps and the wide distribution of their sizes it is necessary to compute correlation functions and power spectra with methods that account for uneven sampling. For direct computation of frequency domain quantities (*e.g.* Fourier transforms, phase and power spectra) the Lomb-Scargle Periodogram (Vanček 1971; Gottlieb et al. 1975; Lomb 1976; Scargle 1982; Scargle 1989), is used here and has been used to study this very data (Donahue and Keil 1995). This appendix

describes the computation of correlation functions for arbitrarily spaced data for their own sakes as well as as an alternative route to these frequency domain quantities.

These computations use the correlation algorithm (Edelson and Krolik 1988) often used in astronomy and well studied in the signal processing literature, under the name of *slotted techniques* (Stoica and Sandgren 2006; Rehfeld et al. 2011). It is an effective way to estimate the auto-correlation function for unevenly spaced time series data such as we have here. Start by constructing bins in the lag variable τ , and then sum the product $x(t_1)x(t_2)$ over all data pairs such that the difference $t_1 - t_2$ lies in each given such bin:

$$\tau_k \leq t_1 - t_2 \leq \tau_k + \Delta\tau, \quad (4)$$

where τ_k denotes the start of bin k and $\Delta\tau$ is the bin width. That is, for data series of the form $X_n = X(t_n)$

$$\rho(\tau_k) = \frac{1}{N_k} \sum_n X_n X_m \quad (5)$$

where the sum is over all pairs n, m such that the corresponding time difference $t_n - t_m$ lies within the bin defined in Eq. (4), and N_k is the number of terms in the sum. It is more usual to write this formula replacing X_n with $X_n - \mu_X$, where μ_X is the mean value of X , either theoretical or empirical. Here we assume an empirical mean has been subtracted. The basic idea is that the average product $x(t_1)x(t_2)$ describes the degree to which values separated by τ are related (large if positively correlated, large and negative if anti-correlated, and small if uncorrelated).

The role of the factor $1/N_k$ is interesting. In estimating correlation functions for evenly spaced data two variants are used in textbooks:

$$\rho(k) = \frac{1}{N} \sum_n X_n X_{n+k} \quad (6)$$

and

$$\rho(k) = \frac{1}{N-k} \sum_n X_n X_{n+k} \quad (7)$$

representing a trade-off favoring small variance (with larger bias) or no bias (with larger variance), respectively. Equation (5) corresponds to equation (7) in that in both cases the denominator in the prefactor is the number of terms contributing to that value of the lag,

so the expression is a true unbiased average. If desired the analog of Equation (6) could be implemented simply by replacing N_k with a constant.

Even though Equation (5) seems a bit abstract it is easily computed in practice. For estimating autocorrelation functions the bins in lag need not be uniform; *e.g.* they could be logarithmic in τ or spaced in any other convenient way. However for computing an estimate of the power spectrum, using the well-known relation that the power spectrum is the Fourier transform of the autocorrelation function, it is most practical to choose uniform bins so that the fast Fourier transform can be used. For evenly spaced data with gaps the binning in τ should correspond to the constant sampling interval. The binned array of lags should extend to the maximum lag supported by the time series, namely the length of the whole observation interval – or the size of a sliding window in the case of time-frequency distributions. These choices were used in all of the power spectrum computations reported in this paper. The sampling is dense enough that there are no empty bins, which would require special attention.

With an algorithm in hand to compute the power spectrum (either the procedure just outlined or the Lomb-Scargle periodogram) it is completely straightforward to compute the time-frequency distribution simply by accumulating a matrix of power spectra of the data points in a sequence of windows slid along the observation interval. The most important parameter is the width of the window. A good choice with the present data was found to be about 0.05 times the whole interval.

REFERENCES

- Arias-Castro, E., Donoho, D., & Huo, X. 2003, IEEE Transactions on Information Theory, 51, 2402
- Babu, P., and Stoica, P. (2010), “Spectral analysis of nonuniformly sampled data - a review,” *Digital Signal Processing*, **20**, 359-378.
- Bai, T. and Sturrock, P. (1993) “Evidence for a fundamental period of the Sun and its relation to the 154 day complex of periodicities,” *Astrophys. J.*, **409**, 476.
- Bertello, L., Pevtsov, A., and Pietarila, A. 2012, ApJ, 761, 11.
- Brown, T. M., Christensen-Dalsgaard, J., Dziembowski, W. A., Goode, P., Gough, D. O., Morrow, C. A. (1989) “Inferring the sun’s internal angular velocity from observed p-mode frequency splittings,” *Astrophysical Journal*, vol. 343, p. 526-546.
- Brevdo, E., and Wu, H.-T. 2011, The Synchrosqueezing Toolbox.
<http://www.math.princeton.edu/~ebrevdo/synsq/>.
- The Synchrosqueezing algorithm: a robust analysis tool for signals with time-varying spectrum, E. Brevdo, N.S. Fu?kar, G. Thakur, and H-T. Wu, 2011.
- Daubechies, I., Lu, J. and Wu, H.-T. 2011, Synchrosqueezed Wavelet Transforms: an Empirical Mode Decomposition-like Tool, *Applied and Computational Harmonic Analysis*, 30, 243
- Domingo, V., G Ermolli, I., Fox, P., Haberreiter, M., Krivova, N., Kopp, G., Schmutz, W., FLomb Solanki, S. K., Spruit, H. C., Unruh,m Y., and Vögler. (2009), “Solar Surface Magnetism and Irradiance on Time Series from Days to the 11-Year Cycle,” , *Space Sci. Rev.*, 145, 337
- Donohue, R., and Keil, S. (1995) The Solar Surface Diffefferential Rotation from Disk-Integrated Chromospheric Fluxes, *Solar Physics*, **159**, 53-62.
- Edelson, R. A. and Krolik, J. H. (1988), *Astrophysical Journal*, 333, 646.

Flandrin, P., Time-Frequency/Time-Scale Analysis (Temps-Frquence) (1999), Academic Press: London

Frasca, A., Frhlich, H.-E., Bonanno, A., Catanzaro, G., Biazzo, K. and Molenda- Zakowicz, J. (2011), “Magnetic activity and differential rotation in the very young star KIC 8429280,” *Astronomy and Astrophysics*, 532, p. A81

Georgieva, K., “Why the Sunspot Cycle is Double Peaked,” *ISRN Astronomy and Astrophysics*, vol. 2011, Article ID 437838

Gonzalez, W., Gonzalez, A., and Tsurutani, B. (1990), “Dual-Peak Solar Cycle Distribution of Intense Geomagnetic Storms,” *Planet. Space Sci.*, **38**, 181-187.

“Optical studies of UHURU sources. XI. A probable period for Scorpius X-1 = V818 Scorpil,” (1975) *Astrophysical Journal Letters*, Vol. 195, p. L33 - L35.

Hall, J. C. (2008). *Stellar Chromospheric Activity*. *Living Reviews in Solar Physics*, 5, 2-+. Retrieved from http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2008LRSP....5....2H

Hathaway, D. (2010), “The Solar Cycle,” *Living Rev. Solar Phys.*, **7**, 1. <http://www.livingreviews.org/lrsp-2010-1>

Hill, M., Hamilton, D., and Krimigis, S. (2009) “Periodicity of 151 days in outer heliospheric anomalous cosmic ray fluxes,” *Journal of Geophysical Research*, **106**, 8315-8322.

Rachel Howe, “Solar Interior Rotation and its Variation,” *Living Rev. Solar Phys.* 6, (2009), 1. URL (cited on jdatej): <http://www.livingreviews.org/lrsp-2009-1>

“An algorithm for optimal partitioning of data on an interval,” Jackson, B., Scargle, J.D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tun Tao Tsai, *IEEE Signal Processing Letters* Vol.12, No. 2, 105- 108

Joshi, B., and Joshi, A. (2005) “Intermediate-Term Periodicities in Soft X-ray Flare Index During Solar Cycles 21, 22, and 23,” *Solar Physics*, **226**, 153-161.

Ca II K-line Monitoring Program, Keil, S., Henry, T., White, D., and Livingston, B., <http://nsosp.nso.edu/data/cak.pdf>. This and other relevant documents,

- including the data analyzed in this paper, are available at the National Solar Observatory web site http://nsosp.nso.edu/data/cak_mon.html, the data being at <ftp://ftp.nso.edu/idl/cak.parameters>.
- Keil, S., Worden, S. P. (1984), Variations in the Solar Calcium K Line 1976-1982, *Astrophysical Journal*, **276**, 766-781.
- Keil, S., Henry, T., and Fleck, B. (1998), NSO/AFRL/Sac Peak K-line Monitoring Program, *Synoptic Solar Physics*, ASP Conference Series, **140**, Balasubramaniam, Harvey and Rabin, eds., also at http://nsosp.nso.edu/data/cak_mon.html.
- Khramova, M., Kononovich, E., and Krasotkin, S. (2002), “Solar cyclicity: fine structure and forecasting,” in *Solar variability: from core to outer frontiers*. The 10th European Solar Physics Meeting, 9 - 14 September 2002, Prague, Czech Republic. Ed. A. Wilson. ESA SP-506, Vol. 1. Noordwijk: ESA Publications Division, ISBN 92-9092-816-6, 2002, p. 145 - 148.
- Livingston, W., Wallace, L., White, O. R., and Giampapa, M. S. (2007), “Sun-as-a-Star Spectrum Variations 1974-2006,” *Ap. J.*, **657**, , 1137-1149.
- Livingston, W., Penn, M. J., and Svalgaard, L. (2012) “Decreasing Sunspot Magnetic Fields Explain Unique 10.7 cm Radio Flux,” *Astrophysical Journal Letters*, 757, L8
- Lomb, N. R. (1976) “Least-squares frequency analysis of unequally spaced data,” *Astrophysics and Space Science*, vol. 39, p. 447
- Lou, Y., Wang, Y., Fan, Z., Wang, S., and Wang, J. (2003) “Periodicities in Solar Coronal Mass Ejections,” *M.N.R.A.S.*, **345**, 809-818.
- Maxxi, M., Niedhöfer, Carpentier, Y., and Ros, E. (2005), “Discovery of Solar Rieger periodicities in another star,” *A.&A.*, **435**, L1-L4.
- Percival, D. B. and Walden A. T. (1993) *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press, Cambridge, UK

- Rehfeld, K., Marwan, N., Heitzig, J. and Kurths, J. (2011) “Comparison of correlation analysis techniques for irregularly sampled time series,” *Nonlinear Processes in Geophysics*, Vol. 18, 389-404
- Rieger, E., Share, G., Forrest, D., Kanbach, G., Reppin, C., and Chupp, E. (1984) “A 154-day periodicity in the occurrence of hard solar flares,” *Nature*, **312**, 623.
- Scargle, J. D. (1982), “Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data,” *Astrophysical Journal*, **263**, 835-853
- Scargle, J. D. (1989), “Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data,” *Astrophysical Journal*, **343**, 874-887.
- Scargle, J. D. (1990), “Studies in astronomical time series analysis. IV - Modeling chaotic and random processes with linear filters”, *Astrophysical Journal*, **359**, 469.
- Scargle, J. 1998 (1998) Studies in astronomical time series analysis. IV - Bayesian Blocks, a new Method to 'Analyze Structure in Photon Counting Data', *Astrophysical Journal*, 504, 405.
- Scargle, J., Norris, J., & Jackson, B. and Chiang, J. (2013) “Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations,” *ApJ* in press and <http://arxiv.org/abs/1207.5578>
- Schrijver, C. J. and Zwaan, C. (2000) *Solar and Stellar Magnetic Activity*, Cambridge University Press.
- Silva-Valio, A., and Lanza, A. F. (2011) “Time evolution and rotation of starspots on CoRoT-2 from the modelling of transit photometry,” *Astronomy and Astrophysics*, 529, A36.
- Stoica, P. and Sandgren, N. (2006) ”Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches,” *Digital Signal Processing*, vol. 16, 712.
- Sturrock, P. (1996) ”A Conjecture Concerning the Rieger and Quasi-Biennial Solar Periodicities,” *astro-ph/9609150*

- The Synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications, Thakur, G., Brevdo, E., Fuckar, N., and Wu, H-T. 2012 <http://arxiv.org/abs/1105.0010>
- Van?ek, P. (1971), “Further Development and Properties of the Spectral Analysis by Least-Squares,” *Astrophysics and Space Science*, Vol. 12, p. 10.
- Way, M.J., P.R. Gazis, and J.D. Scargle (2011), Structure in the 3D galaxy distribution: I. Methods and example results. *Astrophys. J.*, **727**, No. 48,
- White, O., Livingston, W., Keil, S., and Henry, T., Variability of the Solar Ca II K Line over the 22 Year Hale Cycle (1998), *Synoptic Solar Physics*, ASP Conference Series, **140**, Balasubramaniam, Harvey and Rabin, eds., also at http://nsosp.nso.edu/data/cak_mon.html.